GWRBoost:
一种基于局部加性模型与梯度提升的地理加权回归方法

王晗
HanwGeek@gmail.com
导师: 黄舟副教授 & 刘瑜教授

北京大學
PEKING UNIVERSITY

Contents

引言

针对具有 k 个变量的 N 个独立观测, 传统线性回归有:

$$\overbrace{y_i}^{\text{因变量}} = \overbrace{\beta_0 + \sum_{k=1}^{K} \beta_k}^{\text{估计参数}} \underbrace{x_{ik}}_{\text{自变量}} + \underbrace{\varepsilon}_{\text{随机误差项} \sim N(\mu,\, \sigma^2)} \tag{1}$$

采用最小二乘来估计相应系数

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \left[y_i - (\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik}) \right]^2 \tag{2}$$

系数 β_k 代表了对自变量 x_k 和 y 之间关系的可解释量化.

地理加权回归 (Geographically weighted regression, GWR)

[空间异质性] 在地理问题语境下,

变量间关系在研究区域内不一定保持不变. (Goodchild, 2004)

[解决方法] 局部化系数

针对位置 (u_i, v_i) 的特定系数

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \boxed{\beta_k(u_i, v_i)}\, x_{ik} + \varepsilon \tag{3}$$

[优化] 对第 i 个观测使用加权最小二乘

$$\min_{\boldsymbol{\beta}(u_i, v_i)} \sum_{i=n}^{N} \boxed{W[(u_i, v_i), (u_n, v_n)]} \cdot \left[y_n - \left(\beta_0(u_i, v_i) + \sum_{k=1}^{K} \beta_k(u_i, v_i)\, x_{nk} \right) \right]^2 \tag{4}$$

i 处位置 (u_n, v_n) 的空间权重

权重由空间核函数生成 (e.g. bi-square), 生成比普通最小二乘 (OLS) 更好的参数估计

在城市发展, 交通分析, 环境研究等中有大量应用

现存问题

1. 线性模型容易欠拟合 (Schell & Singh, 1997)

- 相较于更复杂模型 (e.g. 决策树, SVM, 神经网络) 性能较差

[解决方法]

- Geographically neural network weighted regression (Du et al., 2020)
- Spatial regression graph convolutional neural networks (Zhu et al., 2021)

但是,

- 复杂模型不能够生成对变量关系的显式解释
  XGBoost + SHAP (Li, 2022)
- 很难使用 AIC 进行评估 (有过拟合风险)

现存问题

2. 加权最小二乘不一定达到全局最优

- 与全局评价指标不一致 (e.g. AIC, RSS)
- 每个观测样本独立地各自优化

需要开发一个模型, 满足

- 有较高模型复杂度, 可以处理大量数据, 避免欠拟合
- 生成变量之间空间变异关系的显式量化
- 使用全局优化函数, 与评价标准一致

GWRBoost

集成学习 & 提升算法

[逐步优化策略] 梯度下降算法

$$\underbrace{\theta_i}_{\text{一般模型参数}} = \theta_{i-1} + \Delta\theta_i = \theta_{i-1} - \lambda \left[\underbrace{\frac{\partial \mathcal{L}(\theta)}{\partial \theta}}_{\text{参数 } \theta \text{ 的梯度}} \right]_{\theta=\theta_{i-1}} \tag{5}$$

$$\underbrace{F_i}_{\text{全局非参数模型 } F_\theta(x)} = F_{i-1} + \Delta F_i = F(x)_{i-1} - \lambda \left[\underbrace{\frac{\partial \mathcal{L}(F)}{\partial F}}_{\text{模型 } F \text{ 的梯度}} \right]_{F=F_{i-1}} \tag{6}$$

[加性模型]

$$\underbrace{f(x)}_{\text{机器学习基模型}} \sim -\lambda \left[\frac{\partial \mathcal{L}(y, F(x))}{\partial F(x)} \right]_{F=F_{i-1}(x)} \Rightarrow \sum f(x) = F(x) \tag{7}$$

集成学习

集合简单 & 较弱的基模型以

- 提高模型复杂度 & 生成更好的结果

[局部] 局部加性模型

$$y_i = F_i(x_i) = \sum_{m=1}^{M} \underbrace{f(x;\beta^m)}_{\text{线性回归}} = \underbrace{\sum_{m=1}^{M} \beta_0^m + \sum_{m=1}^{M}\sum_{k=1}^{K} \beta_k^m(u_i, v_i)\, x_{ik}}_{\text{仍然保持线性形式}} + \varepsilon_i \tag{8}$$

[全局]

$$F(x) = \{F_1, F_2, \ldots, F_N\} \tag{9}$$

[梯度提升优化]

$$\underbrace{f_{\beta^m}(x_i)}_{\text{地理加权回归模型}} \sim \lambda \frac{\partial \mathcal{L}}{\partial F^{m-1}} = \lambda \frac{\partial \left[y - F^{m-1}(x)\right]^2}{\partial F^{m-1}} = \lambda \cdot \frac{1}{N}\sum_{i=1}^{N}\Big[\underbrace{y_i - F^{m-1}(x_i)}_{\text{上一步得出的残差}}\Big] \tag{10}$$

[局部] 局部加性模型

线性回归　　仍然保持线性形式

$$y_i = F_i(x_i) = \sum_{m=1}^{M} f(x;\beta^m) = \sum_{m=1}^{M} \beta_0^m + \sum_{m=1}^{M}\sum_{k=1}^{K} \beta_k^m (u_i, v_i)\, x_{ik} + \varepsilon_i \tag{8}$$

[全局]

$$F(x) = \{F_1, F_2, \ldots, F_N\} \tag{9}$$

[梯度提升优化]

上一步得出的残差

$$f_{\beta^m}(x_i) \sim \lambda \frac{\partial \mathcal{L}}{\partial F^{m-1}} = \lambda \frac{\partial \left[y - F^{m-1}(x)\right]^2}{\partial F^{m-1}} = \lambda \cdot \frac{1}{N} \sum_{i=1}^{N} \left[y_i - F^{m-1}(x_i) \right] \tag{10}$$

地理加权回归模型

以地理加权的方式学习上一步的残差而不是真值

GWRBoost

Algorithm 1: GWRBoost

Data: D =
$\{(X_1, y_1, u_1, v_1), \ldots, (X_N, y_N, u_N, v_N)\}$

Result: Model set $\mathcal{F}^{\mathcal{M}} = \{F_1^M, \ldots, F_N^M\}$

for $n = 1$ **to** N **do**

$\quad \hat{\beta}_n^1 = \arg\min_{\beta_n^1} \frac{1}{2} \sum_{i=1}^{N} w_i(y_i - f_{\beta_n^1}(x_i))^2$

$\quad F_i^1 = f_{\hat{\beta}_n^1}$

end

for $m = 2$ **to** M **do**

$\quad$ **for** $n = 1$ **to** N **do**

$\quad\quad r_n = \lambda \cdot [y_n - F_n^{m-1}(x_n)]$

$\quad\quad \hat{\beta}_n^m = \arg\min_{\beta_n^m} \frac{1}{2} \sum_{i=1}^{N} w_i(r_n - f_{\beta_n^m}(x_i))^2$

$\quad\quad F_n^m = F_n^{m-1} + f_{\hat{\beta}_n^m}$

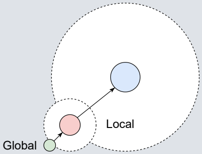
$\quad$ **end**

end

总结

- 初始化一个 GWR
- 收集所有的残差
- 持续训练新的 GWR 以拟合残差

残差传递



通过逐渐收集残差来捕捉全局信息

AIC 计算

$$AIC = -2ln(\underbrace{\hat{\mathcal{L}}}_{\text{似然函数}}) + 2\underbrace{k}_{\text{自由度: 帽子矩阵 } \mathcal{H} \text{ 的迹}} \tag{11}$$

$$\hat{y} = X\hat{\beta} = X\left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}y = \underbrace{\mathcal{H}}_{\text{从真值 } y \text{ 到预测值 } \hat{y} \text{ 的映射}}y \tag{12}$$

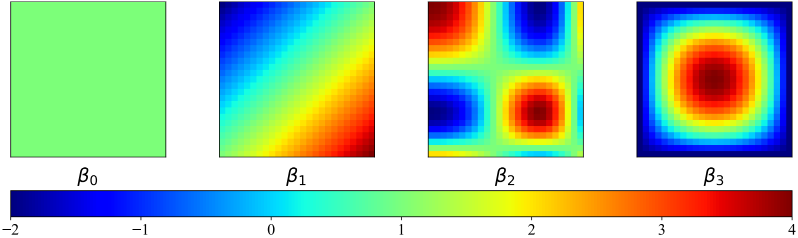
$$\hat{y} = \sum_{m=1}^{M}\hat{y}_m = \sum_{m=1}^{M}\mathcal{H}y_m = \mathcal{H}\sum_{m=1}^{M}(I-\mathcal{H})^{m-1}y_1 = \left\{\underbrace{\mathcal{H}\sum_{m=1}^{M}[\lambda(I-\mathcal{H})]^{m-1}}_{M \text{ 个学习器的帽子矩阵}}\right\}y \tag{13}$$

复杂模型 (e.g. 决策树, 神经网络) 的 AIC 很难度量

实验评估

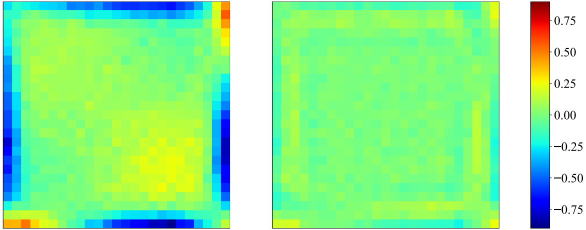
模拟实验

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \tag{14}$$



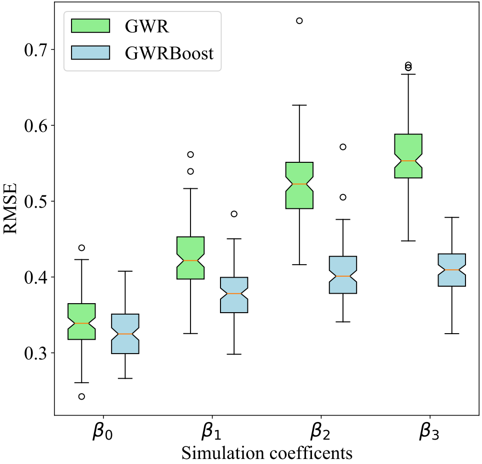
Model	OLS	GWR	**GWRBoost**
RSS	1639.063 ± 72.52	83.900 ± 5.049	**36.797** ± **2.601**
AIC	2385.642 ± 27.65	773.374 ± 36.050	**225.512** ± **42.061**
AICc	2385.739 ± 27.65	839.926 ± 35.383	**274.817** ± **41.207**
R^2	0.072 ± 0.02	0.952 ± 0.003	**0.979** ± **0.002**
Adjusted R^2	0.066 ± 0.02	0.940 ± 0.004	**0.975** ± **0.002**

系数估计误差



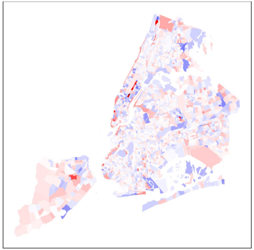
(a) GWR (b) GWRBoost

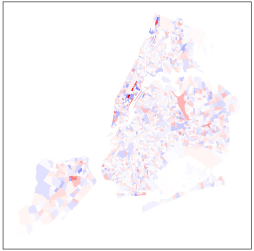
Figure: 残差空间分布

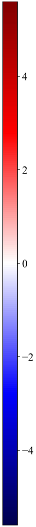


- 降低了边缘误差
- 更低的系数估计 RMSE
- 更低的方差界
- 在异质性强的关系中有更好的性能

案例研究 — NYC 教育数据集


Spatial distribution of GWR residuals


Spatial distribution of GWRBoost residuals



变量	**解释**
因变量	
mean_inc	人均收入
自变量	
sub18	低于 18 岁的人口数
PER_PRV_SC	入学私立学校的学生比例
YOUTH_DROP	16-19 岁辍学青少年比例
HS_DROP	25 岁以上人群辍学比例
COL_DEGREE	25 岁以上至少有一个学士学位比例
SCHOOL_CT	学校数量

Model	OLS	GWR	**GWRBoost**
RSS	982.206	388.626	**261.478**
AIC	4499.669	3168.118	**2289.994**
AICc	4499.720	3315.637	**2437.513**
R^2	0.557	0.825	**0.882**
Adjusted R^2	0.556	0.790	**0.858**
Moran's I	0.333	0.066	**-0.027**

结论

结论

怎样有效？为什么有效？

- 学习残差而不是真值以维护合适的目标函数
- 通过残差传递过程收集全局信息

结论: 我们提出了一个模型能够

- 使用梯度提升算法来优化
- 提高模型复杂度以应用到大规模数据集上
- 被 AIC/AICc 评估
- 保持了生成空间关系显式量化的能力

未来议题

[计算开销] · [带宽选取] · [更多集成学习方法的应用]

感谢老师和同学!
Q & A

References

- Goodchild, M.F., 2004. The validity and usefulness of laws in geographic information science and geography. Annals of the Association of American Geographers, 94 (2), 300–303.
- Du, Z., et al., 2020. Geographically neural network weighted regression for the accurate estimation of spatial non-stationarity. International Journal of Geographical Information Science, 34 (7), 1353–1377.
- Zhu, D., et al., 2021. Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. GeoInformatica, 1–32.
- Li, Z., 2022. Extracting spatial effects from machine learning model using local interpretation method: An example of shap and xgboost. Computers, Environment and Urban Systems, 96, 10184.
- Schell, M.J. and Singh, B., 1997. The reduced monotonic regression method. Journal of the American Statistical Association, 92 (437), 128–135.