

## 空間的クラスタリング構造を取り込んだ回帰分析手法

菅澤 翔之助<sup>1</sup>, 村上 大輔<sup>2</sup>,

<sup>1</sup> 東京大学 空間情報科学研究センター, <sup>2</sup> 統計数理研究所 データ科学研究系

連絡先: <sugasawa@csis.u-tokyo.ac.jp> Web: <https://sites.google.com/view/ssugasawa>

**(1) 動機:** 空間データに対する回帰分析の方法論として、空間加重回帰 (GWR; Geographically Weighted Regression) モデルが標準的なツールとして広く用いられている。この手法は説明変数が被説明変数に与える影響の要因が空間的に変化することを想定し、その影響度を推定することが可能な方法論である。しかし、GWR は数値的に不安定であることが知られており、極端に大きな推定値や見せかけの関係性を誤検出してしまふ可能性がある。

**(2) 方法:** このような問題点を解決するために、本研究では空間的クラスタリング構造 (データが何らかの意味で有限個のグループに分割できること) を想定した新しい空間回帰分析の方法 (SCR; Spatially Clustered Regression) を開発した。提案手法である SCR は、機械学習の分野で広く用いられている K-means クラスタリングと回帰分析手法を融合した方法論である。同一クラスター内では同じ回帰係数の値を持つという制約を課すことで、GWR の問題点として知られている数値的不安定性を解消することが可能になる。さらに、回帰モデルをベースにした空間的クラスタリングにより、解析結果の解釈性も高めることが可能になると考えられる。以上のような利点に加え、SCR は様々な方向への拡張が容易である。例えば、被説明変数に影響与えていないと考えられる説明変数を取り除くこと (変数選択) やより柔軟な回帰モデル (ノンパラメトリック手法) の利用は、提案アルゴリズムの比較的マイナーな修正によって実行可能である。

提案手法である SCR を 2016 年東京 23 区の刑法犯認知件数のデータセットへ適用した。データは 23 区内の町丁目ごとに得られており、欠損部分を取り除いた 2855 地域のデータを解析対象とした。被説明変数として総件数を取り、説明変数として人口密度、昼間人口、外国人比率、失業率、単身世帯率、平均滞在期間をとった。被説明変数が計数データのため、ポアソン回帰モデルをベースにした SCR を適用し、各地域の面積を用いて切片項の調整を行なった。また、比較のためにポアソン回帰モデルを用いた GWR も適用した。

**(3) 結果:** 図 1 は GWR および SCR によって得られた回帰係数 (上から人口密度、昼間人口、失業率に対応する回帰係数) の推定値の空間分布を表している。この結果から明らかなように、代表的な既存

手法である GWR による推定値は空間的に変動が激しく、推定精度・解釈性の両方の観点から好ましい結果とは言い難い。一方で、提案手法の SCR は空間的なクラスタリングの効果により、解釈の容易な推定結果が得られる。今回用いた説明変数の中で、犯罪件数に影響を与える変数を特定するため、正則化を用いた変数選択も考える必要があるが、詳しい結果は当日報告する。

### (4) 使用したデータ

- ・ 筑波大学社会工学コモンズ・データバンク 雨宮 護・小地域時系列犯罪統計データベースの作成 (<https://commons.sk.tsukuba.ac.jp/data>)

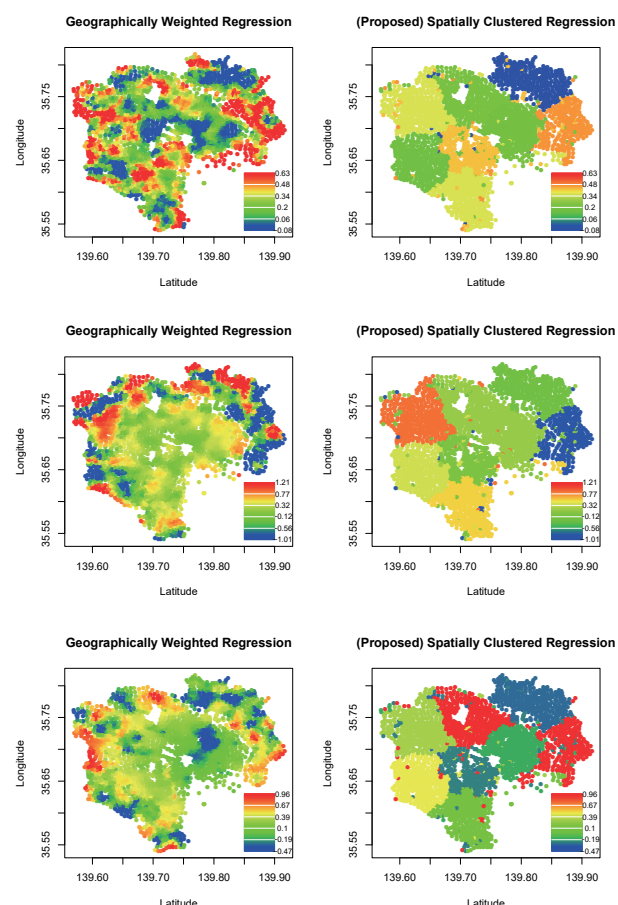


図 1: GWR および SCR (提案手法) による回帰係数の推定値の空間分布 (上から人口密度、昼間人口、失業率に対応する回帰係数に対応する)。