Group lassoに基づく複数階層の領域分割に依存した空間的異質性の抽出

竹本 一至*・井上 亮**

Extraction of Spatial Heterogeneity depending on Multi-level Regional Division by Group Lasso

Kazushi TAKEMOTO*, Ryo INOUE**

Abstract: One approach to analyze spatial heterogeneity is to estimate a model with subregion-specific coefficients that define the minimum spatial scale of analysis. The model estimation with fused lasso penalties can extract a series of subregions whose coefficients are estimated to be the same and grasp the spatial heterogeneity with larger spatial scale than the scale of subregion setting. However, as the fused lasso estimation is computation-intensive, there is a limitation in analyzing the spatial heterogeneity at from detailed to broad spatial resolution. This study proposes a new method that can analyze the scale of spatial heterogeneity by the hierarchical settings of subregions and the group lasso penalties. The experiments with 2-layer hierarchical tessellation verified that the proposed method can extract the scale of spatial heterogeneity.

Keywords: 空間的異質性 (spatial heterogeneity), 空間スケール (spatial scale), 階層性 (hierarchy), スパースモデリング (sparse modeling), group lasso (group lasso)

１．はじめに

近年，行政機関が進めるオープンガバメント政策や民間企業の積極的なデータの公開によって，幅広い種類のデータの入手が可能になった．位置情報を持つ空間データも多種・大量に流通するようになってきており，これらを活用した分析への関心が高まっている．

空間データ分析の一つとして，空間現象の生成要因や過程が場所によって異なるという，空間的異質性の構造把握分析が挙げられる．空間解像度が高いデータが入手可能になっているため，これまでよりも詳細に分析できる環境が整ってきているといえる．

空間上で滑らかに変化する空間可変係数を推定することにより空間的異質性の構造を把握する方法として，地理的加重回帰（例えば，Fotheringham et al., 1997）や固有ベクトル空間フィルタリングに基づく空間可変係数モデル (Griffith, 2008) が提案され，多くの地域分析に活用されてきた．

また，よりオーソドックスな方法として，領域境界で不連続に異なる空間的異質性の分析があり，事前に分析者が分析対象領域を領域分割し，領域毎に設定した係数の推定が行われてきた（例えば，Goodman and Thibodeau, 1998）．この領域境界で不連続な空間的異質性の分析について，スパースモデリングの活用が検討され，井上ら (2020) は fused lasso (Tibshirani et al., 2005)，Inoue et al. (2020) は fused-MCP (Jing et al., 2018) を導入した分析を行っている．これらの分析では，分析対象地域に分析の最小空間単位となる領域分割を事前に設定し，各領域に固有の係数を置いたモデルを立てた上で，隣接領域の係数間の差に対する正則化を行うことにより，係数が同じ値に推定される一連の領域を抽出できる．事前設定の領域よりも大きな空間スケールで発生する空間的異質性を捉えることができる．

しかし，lasso による推定では，計算量のオーダーが正則化条件の数の 3 乗以上 (Tibshirani et al., 2005) になるため，領域が細かく分割されて，隣接する領域の組み合わせ数が多いと実行が難しくなる．当然，各説明変数に領域毎の係数を設定して，説明変数による空間的異質性構造の違いの分析も難しい．

不連続的に起こる空間的異質性は，自治体毎の保育環境など住民サービスの違い，小学校毎の教育サービスの違い，鉄道沿線毎の交通サービスの違い，町名による地域イメージの違いなど，市区町村や小

*　学生会員　東北大学 大学院情報科学研究科（Tohoku University）
〒980-8579　宮城県仙台市青葉区荒巻字青葉 6-6-06　E-mail：kazushi.takemoto.q7@dc.tohoku.ac.jp
**　正会員　東北大学 大学院情報科学研究科（Tohoku University）

学校区，駅勢圏，大字・字，街区など，空間スケールが異なる領域分割単位の影響を受けている可能性がある．そこで，これら複数の領域分割を事前に与えた上で，いずれの分割が空間現象の生成過程に関する空間的異質性の表現に適切かを判断することができれば，細かい領域分割に従って隣接領域の係数間の差に対する正則化を行わずに，詳細なスケールから広域なスケールまでの空間的異質性の分析が可能になると期待される．

そこで，本研究は，複数の領域分割を設定した上で，各場所で生じている空間的異質性の空間分割単位を抽出する手法として，スパースモデリングの 1 種である group lasso に注目し，その適用可能性を検討する．なお，本研究では以後，領域分割は階層的に構成されていることを前提とする．

２．Group lasso を活用した空間的異質性の空間スケール分析の提案

2.1. Group lasso

Group lasso (Yuan and Lin, 2006) は，lasso を拡張したスパースモデリングの一種で，複数の変数をまとめたグループ単位で変数選択ができる．Lasso 推定で利用する L_1 ノルムの代わりに，L_1/L_2 ノルムというグループ毎の正則化を利用する．

ここで，n件のデータをp種類の共変量で表す線形回帰モデルを推定する場合を考える．$n\times1$の被説明変数ベクトルを$\boldsymbol{y}$，$n\times p$行列で表される説明変数を$\mathbf{X}$，$p\times1$のパラメータベクトルを$\boldsymbol{\beta}$とする．さらに，p種類の共変量がK個のグループに分けられ，k番目のグループに対応する共変量がp_k種類の時，$n\times p_k$行列の説明変数を$\boldsymbol{X}_k$，$p_k\times1$のパラメータベクトルを$\boldsymbol{\beta}_k$で表すと，group lasso の推定は式(1)で書け，グループの大きさに対する重みとして$\sqrt{p_k}$が正則化項に含まれる．

$$\min_{\beta}\frac{1}{2}\left\|\boldsymbol{y}-\sum_{k=1}^{K}\boldsymbol{X}_k\boldsymbol{\beta}_k\right\|_2^2+\lambda\sum_{k=1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k\|_2 \qquad (1)$$

ただし，$\|\ \|_2$はユークリッドノルム (L_2 ノルム)を表し，λは正則化項に対する重みを表すハイパーパラメータである．

すべてのグループ内の係数が 1 つなら，group lasso は lasso と等しい．なお，グループは事前に設定する必要がある．

2.2. 異なる空間スケールを表現する階層的な地域構造とグループ設定

空間現象が有する空間的異質性のスケールを把握するため，異なるスケールの階層的な領域分割を設定し，group lasso を用いて分析することを提案する．

階層的な領域分割の設定を説明するため，ここでは上層と下層の 2 階層からなる領域分割を考える．上層の 1 領域は下層の複数領域に分割され，下層のすべての領域が上層のいずれかの領域の構成要素となる．また，1 つの下層領域が複数の上層領域に重なることない．図 1 は上層が 2 領域，下層が 8 領域で構成された 2 階層の領域分割を表す．上層のG_1は下層の$g_{11},g_{12},g_{13},g_{14}$の 4 領域に分割され，$G_2$は$g_{21},g_{22},g_{23},g_{24}$に分割される．

この領域分割設定を利用した，空間現象の空間的異質性が生起する空間スケールの分析を考える．上層・下層の領域毎・共変量毎に係数を設定した回帰モデルを立て，上層の 1 領域を 1 グループ，またその領域を分割した下層領域の集合を 1 グループと設定し，group lasso を用いた推定を提案する．

図 1 では，下層は$g_{11},g_{12},g_{13},g_{14}$の 4 領域と$g_{21},g_{22},g_{23},g_{24}$の 4 領域からなる 2 グループ，上層は$G_1,G_2$をそれぞれ 1 要素からなる 2 グループと考え，合計 4 グループを設定する．

この設定で推定すると，上層の領域毎に，上層・下層のいずれの係数で空間現象を表すべきか選択が行われ，空間的異質性が上層の空間分割が表す大域的なスケールで生じているのか，あるいは，下層の空間分割が表すより局所的なスケールで生じているのかを判別できると期待される．

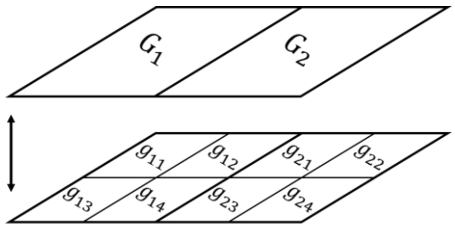


図１ 2 階層構造領域の例

３．実験

階層的な領域分割設定を利用した group lasso による空間的異質性分析の性能確認実験を行う．

3.1. 実験条件

格子領域を設定し，下層の各領域にそれぞれ 20 件のデータを設定する．一部の上層・下層の領域の係数に他の領域とは異なる設定を行い，上層あるいは下層のスケールで発生する空間的異質性を有するシミュレーションデータを生成する．各データの誤差は，平均 0，分散が 1 の正規分布で与える．

正則化項の重みのハイパーパラメータλは，交差検証法で平均二乗誤差が最小となる[A1]ものを選択する．

3.2. 2 階層構造実験

分析対象領域を 2×2 に分割した上層の格子と，上層の各格子を 4×4 に分割し全体領域を 8×8 に分割した下層の格子からなる，階層的な領域分割を設定する (**図 2**)．上層の格子毎に上層・下層それぞれ 4 グループ，計 8 グループを設定した．

本実験では，共変量は定数項のみを考え，分析対象領域全体の定数項α，上層の各領域の係数$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_4)$，下層の各領域の係数$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{64})$を設定する．階層を考慮した分析が行えるかどうか確認するため，「①上層のスケールの空間的異質性を設定した実験」「②下層のスケールの空間的異質性を設定した実験」「③両層のスケールの空間的異質性を設定した実験」の 3 実験を行う．設定はそれぞれ，①上層の係数$\beta_2, \beta_3 = 3$，②下層の係数$\gamma_{18}, \gamma_{23} = 3$，③上層の係数と下層の係数$\beta_3, \gamma_{18}, \gamma_{23} = 3$とする．すべての実験で定数項は$\alpha = 1$とし，設定が与えられなかった領域の係数 β，γは 0 とする．

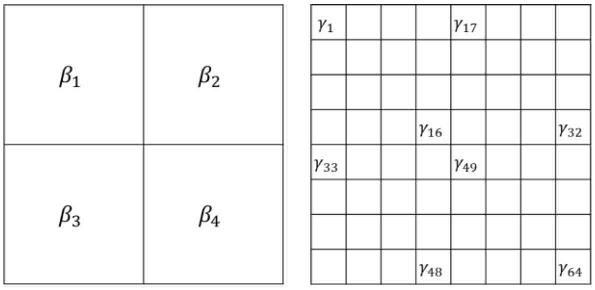
表 1 に各実験の係数推定結果を示す．また，実験②③において，値を与えなかった係数の推定値分布を示すヒストグラムをそれぞれ**図 3・4** に示す．

3.3. 実験結果

実験①では，**表 1** より設定した係数に近い値を推定できたことが確認できる．また，係数を 0 に設定した領域の推定値は，上層の$\widehat{\beta_4} = -0.035$以外はすべての係数が 0 に推定された．上層に設定した空間的異質性は下層領域の係数には表れず，スケールを区別した推定ができていると考えられる．

実験②③では，**表 1** よりγ_{18}, γ_{23}が設定値より小さい値で推定されているが，0 とは異なる大きな値で下層領域の係数を推定できている．この推定値の絶対値が小さくなる傾向は，lasso が有するバイアスによると考えられ，係数が設定されたデータ件数が少ない下層の係数の方が，上層の係数よりもその影響を大きく受けることは，lasso 推定の性質を考えると自然な結果だと言える．また，0 に設定した大半の係数が 0 付近で推定され，両実験とも絶対値最大の推定値が 0.4 付近（**図 3・4**）と，それぞれ下層の係数設定値の 3 よりも十分に小さく，概ね適切な推定結果が得られたと評価できる．

以上の実験から，階層的な領域分割を設定した group lasso による推定により，データが有する空間的異質性のスケールを把握できる可能性があると考えられる．ただし，本稿では，それぞれ一回の実験しか行っておらず，性能評価には至っておらず，更なる検討が必要である．



(1) 上層: $\boldsymbol{\beta}$　　(2) 下層: $\boldsymbol{\gamma}$

図 2　2 階層構造

表 1　2 階層実験の係数推定結果

	実験①		実験②		実験③	
係数	設定	結果	設定	結果	設定	結果
α	1	1.044	1	1.007	1	1.04
β_2	3	2.939	0	0	0	0
β_3	3	2.925	0	0.13	3	2.934
γ_{18}	0	0	3	2.466	3	2.468
γ_{23}	0	0	3	1.919	3	1.921
λ	0.005		0.004		0.004	

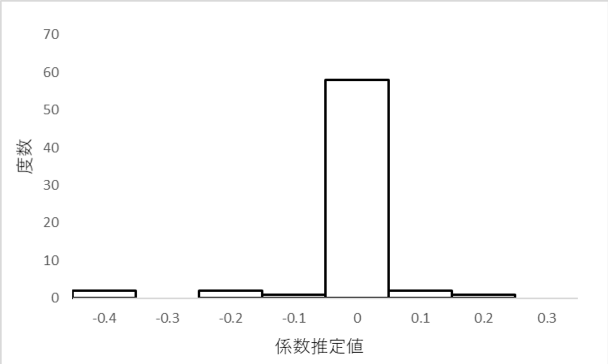


図 3　実験②で 0 に設定した係数の推定値分布

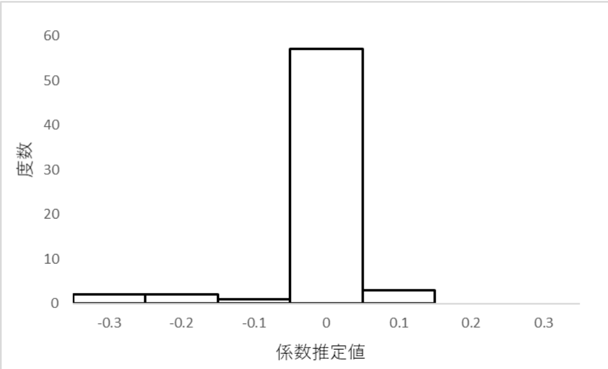


図 4　実験③で 0 に設定した係数の推定値分布

4．おわりに

本研究では，空間現象が有する空間的異質性のスケールを把握する分析方法として，階層的な領域分割に基づき係数設定したモデルを group lasso を適用して推定することを検討した．2 階層の領域分割を用いた実験から，提案分析手法によって空間的異質性のスケールを抽出できる可能性を確認した．

本稿では 2 階層の領域分割を設定した分析に関する検討しか行っていないが，3 階層以上の領域分割を設定した場合に，同様の分析が可能か検討する必要がある．3 階層以上の領域分割を設定した場合，2 階層の場合とは異なって中間層の影響を考慮する必要がある．通常の group lasso の設定では，上層と中間層, 中間層と下層の係数間は変数選択が可能だが，上層と下層の間には直接関係はないため，いずれかのみを用いて表すモデルを推定することは設定できない．複数階層のグループ設定に対応する group lasso の拡張手法が開発されている (Kim and Xing, 2010, Garcia et al., 2014) ため，これらを応用した分析の検討を今後行う.

また，本実験の分析では，説明変数に定数項と領域ダミー変数のみを利用したが，今後，複数の説明変数に対する係数を，領域毎に推定する分析の検討も進めたい．また，領域分割が階層的な構造ではなく，下層領域が複数の上層領域と重なる場合にも分析可能な方法の検討が必要である.

謝辞

本研究は, JSPS 科研費 18H01552 および 21H01447 の助成を受けた.

参考文献

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2003). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.

Garcia, T. P., Muller, S., Carroll, R. J., and Walzem, R. L. (2014) Identification of important regressor groups, subgroups and individuals via regularization methods: application to gut microbiome data. *Bioinformatics*, **30**: 831–837.

Goodman, A. C., and Thibodeau, T. G. (1998) Housing market segmentation. *Journal of Housing Economics*, **7** (2): 121–143.

Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, **40** (11): 2751–2769.

Inoue, R., Ishiyama, R., and Sugiura, A. (2020) Identifying local differences with fused-MCP: An apartment rental market case study on geographical segmentation detection. *Japanese Journal of Statistics and Data Science*, **3**: 183–214.

Jing, B., Yang, G., Yu, X., and Zhang, C. (2018) Fused-MCP with application to signal processing. *Journal of Computational and Graphical Statistics*, **27** (4): 872–886.

Kim, S. and Xing, E. P. (2010) Tree-guided group lasso for multi-task regression with structured sparsity. *27th International Conference on Machine Learning*, 1–14.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005) Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67** (1): 91–108.

Tibshirani, R. (1996) Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **58** (1): 267–288.

Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, **68** (1):49-67

井上 亮, 石山 里穂子, 杉浦 綾子 (2020) 東京都区部の賃貸マンション市場の地理的分割の実態把握－スパースモデリングによるアプローチ－. 土木学会論文集 D3 (土木計画学), **76**(3): 251–263.