

Fused Lasso を用いた地域分類 ～ マンションの賃料に対する地域効果のモデリング ～

広島大 大石 峰暉 大阪医科大 福井 敬祐
(株) 東京カンテイ 岡村 健介, 伊藤 嘉道 広島大 柳原 宏和

本研究ではマンションの賃料の予測問題を扱う。マンションの賃料のような地域に大きく依存するデータを扱うとき、地域が賃料に与える影響を意味する地域効果を考慮する必要がある。実際にマンションの賃料を予測する場合、地域効果の差を取り除くために解析対象を市や区などに限定してモデリングを行い予測することが多い。賃料予測に関する先行研究としては植杉 (2012) や中村・兼田 (2015) などがあるが、いずれも市に限定して予測を行っている。しかしながら、限定した地域内であっても地域効果の差が存在する場合は推定結果にバイアスが生じてしまう。地域効果の差をなくすために、さらに細分化した地域でモデリングを行うことが考えられるが、細分化しすぎると十分な標本が集まらず推定精度の問題が生じる。一方で、細分化した地域でのモデリングではなく、細分化した地域に対応するダミー変数を導入することも考えられるが、細かく分ければ分けるほど変数が増えてしまい予測精度の問題が生じる。このような地域効果の問題は、細分化された各地域の地域効果を推定し、地域効果が等しい地域をまとめるような地域分類をすることができれば解決が期待できる。本研究では、この地域分類を達成するための Fused Lasso を用いた手法を提案する。

解析対象を m 個に細分化し、その中の地域 j に対して、 n_j を標本数、 \mathbf{y}_j をマンションの賃料、 \mathbf{X}_j を説明変数、 ε_j を誤差とする。このとき、地域 j の賃料を以下のようにモデリングすると、地域効果を表現できる。

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mu_j\mathbf{1}_{n_j} + \varepsilon_j \quad (j = 1, \dots, m).$$

ここで、 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ は全地域共通の回帰係数、 μ_j は地域 j の地域効果である。もし東京 23 区を解析対象とし、その細分化を区の範囲で行うとすると、 j は千代田区や港区を表し、 $m = 23$ となる。 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)'$ や $\boldsymbol{\beta}$ の推定には以下のような拡張された Fused Lasso を用いる。

$$\sum_{j=1}^m (\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mu_j\mathbf{1}_{n_j})'(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mu_j\mathbf{1}_{n_j}) + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^m \sum_{k \in D_j} |\mu_j - \mu_k|. \quad (1)$$

ここで、 λ_1, λ_2 は非負のチューニングパラメータ、 D_j は地域 j の隣接地域の集合である。第 2 項は説明変数の変数選択のための罰則項である。通常回帰問題と同様に、使用する説明変数によって推定結果は変わるため変数選択を考える必要がある。第 3 項は地域分類をするための罰則項である。地域の隣接情報を組み込むことで地域効果が等しい隣接地域を結合することができる。(1) 式の最小化問題は陽に解くことができないため、数値探索アルゴリズムが必要である。オリジナルの Fused Lasso 解を得るためには例えば coordinate descent アルゴリズム (Friedman *et al.*, 2007) が提案されている。本発表では (1) 式を最小化するための Friedman *et al.* (2007) を拡張したアルゴリズムを提案し、それを用いた地域分類の結果を示す。

引用文献

- [1] 植杉 大. (2012). 小地域別地価水準のローカル回帰モデル推定 ～ 埼玉県さいたま市を例として. 摂南経済研究, **2**, 1–20.
- [2] 中村 卓馬・兼田 敏之. (2015). 混合地理的加重回帰モデルを応用した地価形成要因の分析. 日本建築学会技術報告集, **21**, 307–310.
- [3] Friedman, J., Hastie, T. & Höfling, H. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.*, **1**, 302–332.