Variable Selection Method for Nonparametric Varying Coefficient Model via Group-lasso Penalty

大阪医科大学	福井 敬祐
広島大学	大石 峰暉
広島大学	小田 凌也
株式会社東京カンテイ	岡村 健介
株式会社東京カンテイ	伊藤 嘉道
広島大学大	栁原 宏和

マンションの価格のような時間や空間に依存するデータに対して個体の変量である築年や占有面積で回帰を行う場合には，時間や空間ごとにこれらの変量が与える影響(回帰係数)が異なることを考慮する必要がある．このような場合には回帰係数が時空間に依存することを仮定した上で柔軟な推定が可能な Nonparametric Varying Coefficient Model（NVCM）の使用が有用である．今 y_i および $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ik})'$ をそれぞれ個体 i の目的変数，説明変数とする($i = 1, \dots, n$)．このとき, NVCM は y_i と $\boldsymbol{x}_i$ の関係を以下のように記述するモデルである．

$$y_i = \beta_0(\boldsymbol{z}_i) + \sum_{j=1}^{k} \beta_j(\boldsymbol{z}_i) x_{ij} + \varepsilon_i.$$

ここで，$\boldsymbol{z}_i =(z_{i1}, \dots, z_{ip})'$ は個体 i が観測された位置(空間)や時点(時間)についての変数であり，$\beta_j(\boldsymbol{z}_i)$ は $\boldsymbol{z}_i$ に依存して変化する未知の回帰係数を表し，変化係数と呼ばれる．さらに，ε_i は誤差で，$\varepsilon_1, \dots, \varepsilon_n$ は互いに独立に同一の平均 0，分散 σ^2 を持つ確率分布に従うとする. NVCM を使用した解析においては，$\beta_j(\boldsymbol{z}_i)$ を基底関数により近似することで，その特徴を柔軟に推定することができるという利点があるが，その柔軟性の代償に使用するデータのサンプルサイズや変量が増加した場合には変数選択や推定が困難であるという問題点がある．そこで本研究では NVCM に対して group lasso 型の罰則を使用した変数選択・推定方法を提案する．変化係数 $\beta_j(\boldsymbol{z}_i)$ を基底関数により以下の様に近似するとする．

$$\beta_j(\boldsymbol{z}_i) = \sum_{\ell=1}^{p} \boldsymbol{\alpha}_{j\ell}' \boldsymbol{g}_\ell(z_{i\ell}) = \boldsymbol{\alpha}_j' \boldsymbol{g}(\boldsymbol{z}_i)\,.$$

ここで，$\boldsymbol{y} = (y_1, \dots, y_n)'$ とし，$\boldsymbol{b}_i = (x_{i1}\boldsymbol{g}(\boldsymbol{z}_i)', \dots, x_{ik}\boldsymbol{g}(\boldsymbol{z}_i)')'$，$\boldsymbol{B} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)'$，$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \dots, \boldsymbol{\alpha}_k')'$ とする．本発表では，以下の罰則残差平方和により推定を行う．

$$\|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha}\|^2 + \boldsymbol{\alpha}'\boldsymbol{Q}\boldsymbol{\Theta}\boldsymbol{Q}'\boldsymbol{\alpha} + 2\lambda \sum_{j=1}^{k} \sum_{\ell=1}^{p} w_{j\ell} \|\boldsymbol{\alpha}_{j\ell}\|\,.$$

ただし，λ は正則化パラメータ，$\boldsymbol{\Theta}$ は対角成分に平滑化パラメータを持つ対角行列，$w_{j\ell}$ は adaptive lasso のための重み，$\boldsymbol{Q}$ は $\boldsymbol{B}$ と基底関数に対応する平滑化行列に依存する直交行列である．上記の罰則付き残差に基づく推定により，$x_{i1}, \dots, x_{ik}$ の変数選択だけでなく，変化係数 $\beta_j(\boldsymbol{z}_i)$ にどの $z_{i1}, \dots, z_{ip}$ が必要かという選択も行うことができる．また，平滑化パラメータの最適化には, Yanagihara（2012）で提案された繰り返し計算を必要としない方法を用いて推定にかかる時間の短縮を図る．発表当日は推定方法の詳細と不動産データへの適用例を紹介する．

引用文献

[1] Yanagihara, H. (2012). A non-iterative optimization method for smoothness in penalized spline regression. *Statist. Comput.*, **22**, 527-544.