

主効果と交互作用の関係を考慮したベイジアン Lasso によるベイズ変数選択

ヤンセンファーマ株式会社 野口英久
東京理科大学 尾島善一
東京理科大学 安井清一

1. 導入と背景

変数選択の目的は、目的変数に影響を与える説明変数を特定することである。実験計画において、変数選択はスクリーニング実験で行われ、主効果と交互作用の特定が目的とされる。スクリーニング実験における変数選択では、主効果と交互作用の関係について Effect Heredity Principle 「交互作用が効果を持つには、少なくとも一つの対応する主効果が効果を持つべきである」という原則に基づいて変数が選択される。Effect Heredity Principle は Strong Heredity Principle 「交互作用に効果がある場合、対応する主効果すべてに効果がある」と Weak Heredity Principle 「交互作用に効果がある場合、対応する主効果すべて、もしくはいずれかに効果がある」に分類できる。

変数選択手法 Lasso(Tibshirani (1996))によって得られる回帰係数の推定値について、回帰パラメータの事前分布にラプラス分布を仮定した場合における事後分布のモードとして解釈する Bayesian Lasso が Park and Casella (2008)によって提案された。しかし、事後分布のモードに基づいた変数の効果の有無の判断は容易でない場合が起こり得る。

Chipman *et al* (1997)や Kuo and Mallick (1998)によるベイズ変数選択では各変数の効果の有無を 0 又は 1 で表すパラメータを導入することにより、変数選択の結果解釈を容易にしている。

3. 目的

本研究では Bayesian Lasso をベイズ変数選択の手法へ適用し、主効果と交互作用の関係を考慮した変数選択手法を提案する。提案された手法をブラケットバーマン計画を用いたスクリーニング実験によって得られたデータに適用し、変数選択を行う。

4. モデル

本研究では Kuo and Mallick (1998)による、変数の効果の有無を 0 又は 1 の値で表す $p \times 1$ の指示パラメータ δ を導入した以下の線形モデルを用いる。

$$Y = XD\beta + \epsilon \quad \epsilon \sim N(0, \sigma^2)$$

Y : n 個の応答変数、 X : $n \times p$ 行列、 D : 対角要素が δ の $p \times p$ 行列、 β : $p \times 1$ の回帰ベクトル、 ϵ : $n \times 1$ の誤差ベクトル

回帰パラメータ β の事前分布はラプラス分布に従い、指示パラメータ δ はベルヌーイ分布に従う。

また、本研究では主効果と交互作用の関係を表す Effect Heredity Principle (Strong Heredity Principle 及び Weak Heredity Principle) を以下のように、指示パラメータの変形によって考慮する。

$$f_{Strong}(x) = \sum_{i=1}^p \beta_i \delta_i x_i + \sum_{i < j} \sum_{j=2}^p \beta_{ij} \delta_{ij} \delta_j x_i x_j$$

$$f_{Weak}(x) = \sum_{i=1}^p \beta_i \delta_i x_i + \sum_{i < j} \sum_{j=2}^p \beta_{ij} \delta_{ij} (1 - (1 - \delta_i)(1 - \delta_j)) x_i x_j$$

5. まとめ

本研究ではベイジアン Lasso に対し、ベイズ変数選択における効果の有無を表すパラメータを導入し、主効果と交互作用の関係を考慮した変数選択手法を提案する。また、Kuo and Mallick (1998)によるベイズ変数選択手法、ベイジアン Lasso 及び提案法を、代表的なスクリーニング計画の一つであるブラケットバーマン計画によって得られたデータに用いて変数選択を行い、それぞれの手法によって得られた結果の比較と考察を行う。

参考文献

- Chipman H, Hamada H, Wu CFJ. (1997), "A Bayesian variable selection approach for analyzing designed experiments with complex aliasing," *Technometrics*, 39:372–381.
- Kuo, L. and Mallick, B.(1998). "Variable Selection for Regression Models," *The Indian Journal of Statistics*. 60: 65-81.
- Park, T., and Casella, G. (2008) "The Bayesian lasso," *J. Am. Stat. Assoc.*, 103, 681-686.
- Tibshirani, R. (1996) "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B*, 58, 267-288.