Clustered lasso を用いたスパース共クラスタリング

# Sparse biclustering via Clustered lasso

数学専攻　上田 朋矢

UEDA, Tomoya

## 1　はじめに

近年, IoT やデータ収集蓄積技術の発展に伴い, SNS データや購買ログデータなどに代表される, 人と人, 人と物などの繋がりを表すデータが膨大に蓄積されている．そのようなデータに対する最も基本的な分析処理として, 類似したデータをグルーピングするクラスター分析がある．購買ログデータが与えられたとき, 通常のクラスタリング手法であれば, 顧客のクラスタリングもしくは商品のクラスタリングを個別に行う．

一方, 本研究の主題である共クラスタリングは, 顧客と商品を同時にクラスタリングする技術である．これにより「もっとも売れる顧客層と商品群」のような有益な情報を持つ部分を発見することができる．共クラスタリングにはベイズアプローチを用いた方法 (Nowicki and Snijders (2001)) や本研究が該当する $k$-means 法を拡張した方法がある．Nowicki and Snijders (2001) は, 非対称関係データにも適応可能な共クラスタリングとして確率的ブロックモデルを提案した．また, Kemp, Tenenbaum, Griffiths, Yamada and Ueda (2006) は, 確率的ブロックモデルを拡張し, 潜在するクラスター数を自動的に決定できる無限関係モデルを提案した．

共クラスタリングにおいて, 行のクラスター数を $K$, 列のクラスター数を $R$ とすると, 合わせて $KR$ 個という多数のクラスターができる．そこで Tan and Witten (2014) は, 中心化後のデータ行列 $X$ におけるクラスター平均の $L_1$ 正則化推定として共クラスタリングをとらえ, いくつかのクラスターの平均が 0 であると推定する方法を提案した．しかし, 平均 0 のクラスター以外にも平均が同一のクラスターは存在しうるが, Tan and Witten の方法では全て別のクラスターとして扱われてしまう．そこで本研究では, Clustered lasso (She (2010)) の罰則項を加えることにより, 平均 0 のクラスターを推定するのみならず, 平均が近いクラスターを同一のクラスターと扱うスパース共クラスタリング法を提案する．

## 2　スパース共クラスタリング

いま, 各個体の特徴を捉える $p$ 個の変数を $\boldsymbol{x} = (x_1, x_2, \dots, x_p)^T$ とする．この $p$ 個の変数に関して観測された $n$ 個の $p$ 次元データベクトルを用いて, 行列 $X$ を $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$ とする．$p$ 個の特徴は $R$ 個の重複しないクラスター $D_1, \dots, D_R$ のいずれかに属し, $n$ 個の観測値は $K$ 個の重複しないクラスター $C_1, \dots, C_K$ のいずれかに属するとする．データ行列 $X$ が中心化されている場合, いくつかの共クラスターの平均の推定値はほぼ 0 である可能性がある．この場合, これらの推定値を正確に 0 と推定することで, 解釈可能性の向上と結果として得られる共クラスターの分散を低減することが出来る．Tan and Witten (2014) で提案された手法は $k$-means 法をスパースで対称にしたものであり, 行列の要素は共クラスター固有の平均と分散をもつ正規分布に従って分布していると仮定し, 対応する対数尤度を最大化することによって共クラスタリングを実行する．

Tan and Witten (2014) のスパース共クラスタリング (sparseBC) では, 中心化されたデータ行列 $X$ に対し

$$\underset{C_1,\dots,C_K,D_1,\dots,D_R,\mu\in\mathbb{R}^{K\times R}}{\text{minimize}} \left\{ \sum_{k=1}^{K}\sum_{r=1}^{R}\sum_{i\in C_k}\sum_{j\in D_r}(X_{ij}-\mu_{kr})^2 + \lambda_1 \sum_{k=1}^{K}\sum_{r=1}^{R}|\mu_{kr}| \right\} \tag{1}$$

を解くことにより, これらの推定を行う. ここで $\lambda$ は正則化パラメータであり, $\lambda$ が増加すると $\mu_{kr}=0$ と推定される $\mu_{kr}$ の数が増加する. この最適化問題は soft-thresholding により表現でき, 繰り返しを行うアルゴリズムによって解くことができる.

## 3 Clustered lasso による共クラスタリング

前節のスパース共クラスタリング法では, いくつかのクラスターの平均を 0 であると推定し同一クラスターとして扱うことにより, 解釈可能性が上がると考えられる. さらに本節では, 0 ではなくても, クラスター平均の非常に近いクラスターを同一クラスターと捉えることができるようにスパース共クラスタリングを拡張する.

前節の目的関数に Clustered lasso（異なるパラメータの差の $L_1$ ノルム正則化）型の制約を加え

$$\underset{C_1,\ldots,C_K,D_1,\ldots,D_R,\mu\in\mathbb{R}^{K\times R}}{\text{minimize}}\left\{\sum_{k=1}^{K}\sum_{r=1}^{R}\sum_{i\in C_k}\sum_{j\in D_r}(X_{ij}-\mu_{kr})^2+\lambda_1\sum_{k=1}^{K}\sum_{r=1}^{R}|\mu_{kr}|\right.$$
$$\left.+\frac{\lambda_2}{2}\sum_{k=1}^{K}\sum_{r=1}^{R}\sum_{i=1}^{K}\sum_{j=1}^{R}|\mu_{kr}-\mu_{ij}|\right\} \tag{2}$$

という最適化問題を考える. ここで $\lambda_1,\lambda_2$ は正則化パラメータであり, $\lambda_1$ が増加すると $\mu_{kr}=0$ と推定される $\mu_{kr}$ の数が増加し, $\lambda_2$ が増加すると $\mu_{kr}=\mu_{k'r'}$ と推定される $\mu_{kr}$ の数が増加する. これにより, いくつかの平均値を 0 と推定しつつ, 値の近いいくつかの平均値を等しく推定する.

この最適化問題を以下のように交互最適化により解くことを考える. いま, vec 作用素を用いて, $\boldsymbol{X}=\mathrm{vec}(X)=(x_{11},x_{12},\ldots,x_{np})^T$ とする. また $\boldsymbol{\mu}=(\mu_{11},\mu_{12},\ldots,\mu_{KR})^T$ とする. さらに, $x_{ij}=\tilde{x}_{s_{ij}}$, $s_{ij}=(i-1)p+j$, $\mu_{kr}=\tilde{\mu}_{t_{kr}}$, $t_{kr}=(k-1)R+r$, $C_k\cap D_r=C\cap D_{t_{kr}}$, $(s_{ij}=1,\ldots,np)$, $(t_{kr}=1,\ldots,KR)$ とする. $P$ は $np\times KR$ 型計画行列で

$$P_{i'j'}=\begin{cases}1 & (i'=s_{ij},\ j'=t_{kr},\ \tilde{x}_{s_{ij}}\in C\cap D_{t_{kr}})\\ 0 & (\text{otherwise})\end{cases}$$

とする. $\boldsymbol{\varepsilon}$ は $np$ 次元の誤差ベクトルである. この下で $\boldsymbol{\mu}$ を推定する問題は,

$$\boldsymbol{X}=P\boldsymbol{\mu}+\boldsymbol{\varepsilon} \tag{3}$$

という回帰問題に帰着できる. 式 (4.2) を Clustered lasso を用いてスパース推定する. 最小化するのは

$$\frac{1}{2n}||\boldsymbol{X}-P\boldsymbol{\mu}||_2^2+\lambda||D\boldsymbol{\mu}||_1 \tag{4}$$

である. ここで $D=\frac{1}{\lambda}(\lambda_1 I_p+\lambda_2\tilde{D})$, $\tilde{D}$ は ${}_{KR}\mathrm{C}_2\times KR$ 型差分行列であり, 1 から $KR$ までの整数を用いた ${}_{KR}\mathrm{C}_2$ 個の組み合わせを $(a,b)_c$, $(1\le a,b\le KR)$, $(c=1,\ldots,{}_{KR}\mathrm{C}_2)$, とすると

$$\tilde{D}_{ij}=\begin{cases}1 & (i=c,\ j=a)\\ -1 & (i=c,\ j=b)\\ 0 & (\text{otherwise})\end{cases}$$

である. $L_1$ 正則化を用いたスパース共クラスタリングと同様の手順により式 (2) の最適解を求める．Clustered lasso によるスパース共クラスタリングでは，$\mu_{kr}$ の推定を式 (4) のように Generalized lasso の形で表すことができ，その最小化は座標降下法などにより行うことが可能である．つまり，平均の縮小推定（Generalized lasso）と，距離が最小となるクラスターへの割り当てを，行と列とで順に繰り返していけばよい．ただし，2 節

**Algorithm 1** Clustered lasso によるスパース共クラスタリング

1: 中心化したデータ行列 $X$ の列と行それぞれに $k$-means クラスタリングを行い，$C_1, \dots, C_K$ 及び $D_1, \dots, D_R$ に初期値を与える．

2: 収束するまで手順 3 から手順 6 を繰り返す．

3: $C_1, \dots, C_K$ 及び $D_1, \dots, D_R$ を固定し, (4.3) 式 の最小化により, $\mu_{kr}$ を求める．

4: $D_1, \dots, D_R$ 及び $\mu_{kr}$ を固定し, $i$ 番目の観測値を $\sum_{r=1}^{R} \sum_{j \in D_r} (X_{ij} - \mu_{kr})^2$ を最小とする行クラスター $C_k$ に割り当てる．

5: 手順 3 を行う．

6: $C_1, \dots, C_K$ 及び $\mu_{kr}$ を固定し, $j$ 番目の特徴を $\sum_{k=1}^{K} \sum_{i \in C_k} (X_{ij} - \mu_{kr})^2$ を最小とする列クラスター $D_r$ に割り当てる．

の スパース共クラスタリングと異なり，Clustered lasso の制約によって行と列のクラスターが互いに影響し合うため，平均の縮小推定と行のクラスターへの割り当て，平均の縮小推定と列のクラスターへの割り当て，というように，平均の縮小推定をクラスター割り当ての前に毎回行うべきである．このアルゴリズムをまとめたものが Algorithm 1 である．

このアルゴリズムにより式 (2) の最適化に関する局所解が得られると考えられる．ただし，ここで，$\lambda_1, \lambda_2$ の値は，アルゴリズムを通じて固定しておく必要がある．

# 4 数値例

この節では，提案手法 (CLsparseBC) と先行研究である Tan and Witten (2014) の方法 (sparseBC) を，Toy example により比較する．

正規分布から $120 \times 60$ 型データ行列を以下のように発生させる．行のクラスター数 $K$ と列のクラスター数 $R$ は既知であり，それぞれ 3 であるとする．各クラスターに含まれる要素の数は等しく $40 \times 20$ であるとする．各クラスターの平均は表 1 の通りであるとし，それぞれの平均 $\mu_{kr}$ に対して $X_{ij}$ は正規分布 $N(\mu_{kr}, 4^2)$ により発生させる．ここで, $X_{ij}$ の行クラスターが $k$, 列クラスターが $r$ であるとする．これをまとめたものがデータ行列 $X$ である．

## 4.1 実行結果

表 3 は sparseBC, 表 4 は CLsparseBC を実行した結果である．正則化パラメータは sparseBC では Tan and Witten (2014) に従い $\lambda = \sqrt{60 \log 120}$ とし，CLsparseBC では $\lambda_1 = \lambda_2 = \sqrt{60 \log 120}$ とした．図から，どちらの方法においても真の構造にある程度近い結果が得られていることが見てとれる．数値を詳しく見てみると，今回の場合 sparseBC では正則化の効果が見られず，正確に 0 と推定しているクラスターはないことがわかる．推定された平均の値は比較的真値に近いが，$(3, 2)$ クラスターのみ真値とかなり異なる値として推定されてしまっている．一方，CLsparseBC を用いた場合，真の平均が 0 および $-1.5$ のクラスターはうまく併合できており，かつ真値に近い推定値が得られている．一方で，真の平均 2 のクラスターはうまく併合することができなかった．

表 1　Mean Matrix

| | [,1] | [,2] | [,3] |
|---|---|---|---|
| [1,] | -1.5 | 2 | 0 |
| [2,] | 3.5 | -1.5 | 0.5 |
| [3,] | 2 | 0 | 2 |

表 2　真のクラスター構造の標本平均行列

| | [,1] | [,2] | [,3] |
|---|---|---|---|
| [1,] | -1.573 | 1.952 | -0.02 |
| [2,] | 3.814 | -1.895 | 0.222 |
| [3,] | 1.909 | -0.266 | 2.082 |

表 3　sparseBC で推定した各クラスターの平均値

| | [,1] | [,2] | [,3] |
|---|---|---|---|
| [1,] | -1.866 | 1.866 | 0.012 |
| [2,] | 3.849 | -1.517 | 0.777 |
| [3,] | 1.928 | -0.243 | 2.102 |

表 4　CLsparseBC で推定した各クラスターの平均値

| | [,1] | [,2] | [,3] |
|---|---|---|---|
| [1,] | -1.453 | 1.782 | 0.036 |
| [2,] | 3.451 | -1.453 | 1.005 |
| [3,] | 1.602 | 0.036 | 1.976 |

## 5　まとめ

本稿では, 一般的な共クラスタリング手法を紹介し, スパース共クラスタリングについて記述した. さらに, 既存のスパース共クラスタリングを拡張し, Clustered lasso 型の制約を加えることによって, 平均が近いクラスターを同一のクラスターと扱うスパース共クラスタリング法を提案した.

クラスタリングを実行するアルゴリズムは sparseBC の手順で行い, クラスター平均 $\mu_{kr}$ の推定に Generalized lasso を用いた. チューニングパラメータ $K, R$ は sparseBC で紹介された bi-cross-validation を用い, 正則化パラメータ $\lambda_1, \lambda_2$ についても同様に BIC による方法を提案した. 最後に, 数値例により, 平均値が等しいクラスターを発見することができた.

今後の課題として, 最適な $K, R$ を選択する方法の検討が挙げられる. 例えば, $K, R$ を十分に大きくとり正則化パラメータの調整によって小さくし, 調整する方法が考えられる.

## 参考文献


[1] Kemp, C. Tenenbaum, J. B. Griffiths, T. L. Yamada, T. and Ueda, N. (2006). Leanig systems of concepts with an infinite relational model. In *Proccedings of the 21st National Conference on Artificial Intelligence (AAAI)*, pp.381-388.

[2] Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. In *Jounal of the American Statistical Association (JASA)*, 96(455), pp.1077-1087.

[3] She, Y. (2010). Sparse regression with exact clustering. *Electronic Journal of Statistics*, 4, 1055-1096.

[4] Tan, K. M. and Witten, D. M. (2014). Sparse Biclustering of Transposable Data. *Journal of Computational and Graphical Statistics*, 23(4):985–1008.

[5] Tibshirani, R. and Taylor, J. (2010). The solution path of the generalized lasso. *Ann. Statist.* 39, 1335–1371.