正則化法によるカーネル正準相関分析

Kernel canonical correlation analysis via regularization methods

数学専攻　竹森 悠渡

TAKEMORI, Yuuto

1　はじめに

正準相関分析 (Canonical Correlation Analysis) は, Hotelling (1936) によって提案された 2 種類の観測項目群の相関関係を分析するときに用いられる多変量解析手法の 1 つである. 近年, 言語探索, ゲノムデータ解析, 機械学習など幅広い分野で用いられている. 正準相関分析は, 各項目群のデータを線形射影したときの関連性の程度を基準とおく手法であり, 一般化固有値問題に帰着できる.

正準相関分析はデータが高次元の場合, 共分散行列が特異となり, 一般化固有値問題を解くことが難しい. この問題に対して Hardoon *et al.* (2004) は正則化法を用いた正準相関分析を提案した. 正則化法を適用するに当たって, 正則化パラメータ, あるいは調整パラメータの値によって推定したモデルが変化するので, どのように決定するかが重要となる. さらに高次元データを分析の対象とするときは, 正準相関分析に基づく分析結果に不要な情報が過剰に取り込まれることから, 結果の解釈が困難である. この問題に対して回帰分析における回帰係数の一部を 0 に推定するスパースモデリングを正準相関分析に用いることが Witten *et al.* (2011) や Chu *et al.* (2013) によって提案されている.

非線形構造を持つデータに対しては, 線形結合に基づく正準相関分析は有効に機能しない. 近年のデータ解析において, テキスト文書, 画像, マイクロアレイデータなど非線形関係を検討することは重要である. 赤穂 (2000) は非線形構造をもつデータの分析に対してカーネル正準相関分析を提案した. また Chu *et al.* (2013) ではカーネル正準相関分析にスパース性を持たせることを目的とした手法が提案されている.

本論文では, 正準相関分析に関して起因する正則化推定法, スパースモデリング, カーネル法などについて, 理論的に定式化し, 問題点を考察するとともに, その解決法について検討を行った.

2　カーネル正準相関分析

正準相関分析は, 2 種類の観測項目群の相関関係を分析するときに用いられる手法の 1 つである. 正準相関分析では, 各項目群のデータを線形射影したときの関連性の程度を基準とおく手法であり, 2 つの確率変数ベクトルに線形射影を施した変数間の相関係数が最大になるように線形結合のパラメータを推定する.

正準相関分析は, 線形構造を内包しているデータには有効に機能するが, データが非線形構造を有している場合には線形結合に基づく正準相関分析は有効に働かない. これに対して非線形構造を持つデータに対する正準相関分析としてカーネル法を用いた正準相関分析がある.

p 次元データベクトルを $\boldsymbol{x}$, q 次元データベクトルを $\boldsymbol{y}$ とし, それぞれのデータベクトルに関して観測された n 組のデータを $\{\boldsymbol{x}_i, \boldsymbol{y}_i; i = 1, \ldots, n\}$ とする. また, 高次元空間への写像をそれぞれ $\boldsymbol{\phi}_x : \boldsymbol{x}_i \to \boldsymbol{\phi}_x(\boldsymbol{x}_i) \in \mathbb{R}^r (r \gg p)$, $\boldsymbol{\phi}_y : \boldsymbol{y}_i \to \boldsymbol{\phi}_y(\boldsymbol{y}_i) \in \mathbb{R}^t (t \gg q)$ とする. すなわち

$$\boldsymbol{\phi}_x(\boldsymbol{x}_i) = (\phi_{x_1}(\boldsymbol{x}_i), \ldots, \phi_{x_r}(\boldsymbol{x}_i)), \qquad \boldsymbol{\phi}_y(\boldsymbol{y}_i) = (\phi_{y_1}(\boldsymbol{y}_i), \ldots, \phi_{y_t}(\boldsymbol{y}_i))$$

と定義する. ベクトルの各成分は, 各々 p, q 変数実数値関数である. このとき, 高次元空間におけるデータ行列を $X_\phi = (\boldsymbol{\phi}_x(\boldsymbol{x}_1), \ldots, \boldsymbol{\phi}_x(\boldsymbol{x}_n))^T$, $Y_\phi = (\boldsymbol{\phi}_y(\boldsymbol{y}_1), \ldots, \boldsymbol{\phi}_y(\boldsymbol{y}_n))^T$ とする. X_ϕ, Y_ϕ の標本分散, 標本共分散行列を S^ϕ_{xx}, S^ϕ_{yy}, S^ϕ_{xy} とする. X_ϕ, Y_ϕ の 1 次結合 $\boldsymbol{u}^\phi$, $\boldsymbol{v}^\phi$ を係数ベクトル $\boldsymbol{c} = (c_1, \ldots, c_r)^T$, $\boldsymbol{d} = (d_1, \ldots, d_t)^T$ を用いて

$$\boldsymbol{u}^\phi = X_\phi \boldsymbol{c}, \qquad \boldsymbol{v}^\phi = Y_\phi \boldsymbol{d}$$

と表すことができる. $\boldsymbol{u}^\phi$ の標本分散 S^ϕ_u は $S^\phi_u = \boldsymbol{c}^T S^\phi_{xx} \boldsymbol{c}$, $\boldsymbol{v}^\phi$ の標本分散 S^ϕ_v は $\boldsymbol{d}^T S^\phi_{yy} \boldsymbol{d}$ である. また, $\boldsymbol{u}^\phi$, $\boldsymbol{v}^\phi$ 間の標

本共分散は $S_{uv}^{\phi} = \boldsymbol{c}^T S_{xy}^{\phi} \boldsymbol{d}$ であり, $\boldsymbol{u}^{\phi}$, $\boldsymbol{v}^{\phi}$ 間の相関係数 ρ_{ϕ} は次式で与えられる.

$$\rho_{\phi} = \frac{S_{uv}^{\phi}}{\sqrt{S_u^{\phi}}\sqrt{S_v^{\phi}}} = \frac{\boldsymbol{c}^T S_{xy}^{\phi} \boldsymbol{d}}{\sqrt{\boldsymbol{c}^T S_{xx}^{\phi} \boldsymbol{c} \boldsymbol{d}^T S_{yy}^{\phi} \boldsymbol{d}}}.$$

相関係数 ρ_{ϕ} が最大になるパラメータを求めたいが, データを高次元空間に写像したことによって解を求めることは困難である.

この問題に対してカーネル法を用いて克服する. 係数ベクトル $\boldsymbol{c}$, $\boldsymbol{d}$ が次の形で表現できることを用いる.

$$\boldsymbol{c} = X_{\phi}^T \boldsymbol{\alpha}, \quad \boldsymbol{d} = Y_{\phi}^T \boldsymbol{\beta}.$$

ここで, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^T$ はパラメータである. 高次元空間に写像されたベクトル $\boldsymbol{\phi}_x(\boldsymbol{x})$ と $\boldsymbol{\phi}_y(\boldsymbol{y})$ の内積計算を次のようにカーネル関数で置き換えることにより対処していく.

$$k_x(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\phi}_x(\boldsymbol{x_i})^T \boldsymbol{\phi}_x(\boldsymbol{x_j}), \quad k_y(\boldsymbol{y}_i, \boldsymbol{y}_j) = \boldsymbol{\phi}_y(\boldsymbol{y_i})^T \boldsymbol{\phi}_y(\boldsymbol{y_j}).$$

このとき, $\boldsymbol{u}^{\phi}, \boldsymbol{v}^{\phi}$ はパラメータ $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^T$ を用いて次のように表すことができる.

$$\boldsymbol{u}^{\phi} = X_{\phi} \boldsymbol{c} = X_{\phi} X_{\phi}^T \boldsymbol{\alpha} = K_x \boldsymbol{\alpha}, \quad \boldsymbol{v}^{\phi} = Y_{\phi} \boldsymbol{d} = Y_{\phi} Y_{\phi}^T \boldsymbol{\beta} = K_y \boldsymbol{\beta}.$$

K_x, K_y は $k_x(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $k_y(\boldsymbol{y}_i, \boldsymbol{y}_j)$ を成分にもつ $n \times n$ カーネル行列である. カーネル行列を用いると相関係数 ρ_{ϕ} は次のように表すことができる.

$$\rho_{\phi} = \frac{\boldsymbol{\alpha}^T K_x K_y \boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}^T K_x^2 \boldsymbol{\alpha} \boldsymbol{\beta}^T K_y^2 \boldsymbol{\beta}}}.$$

パラメータ $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ に正の数を掛けても相関係数 ρ の値は変わらないので , $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ はこの意味において不定性を有する. 相関係数 ρ_{ϕ} を最大とする係数ベクトル $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ の推定は標本分散の値を 1 とする $\boldsymbol{\alpha}^T K_x \boldsymbol{\alpha} = \boldsymbol{\beta}^T K_y \boldsymbol{\beta} = 1$ の制約条件を課した, 次の最大化問題を解くことになる.

$$\arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{\alpha}^T K_x K_y \boldsymbol{\beta}, \quad \text{subject to } \boldsymbol{\alpha}^T K_x^2 \boldsymbol{\alpha} = \boldsymbol{\beta}^T K_y^2 \boldsymbol{\beta} = 1. \tag{2.1}$$

(**??**) 式の最大化問題は過適合を引き起こしてしまい推定できない. この問題に対して Malte, Kuss., and Thore, Graepel. (2003) では次のように制約を変更した最適化問題に帰着できる.

$$\arg\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{\alpha}^T K_x K_y \boldsymbol{\beta}, \quad \text{subject to } \boldsymbol{\alpha}^T (K_x^2 + \gamma_x I_n) \boldsymbol{\alpha} = \boldsymbol{\beta}^T (K_y^2 + \gamma_y I_n) \boldsymbol{\beta} = 1. \tag{2.2}$$

ここで, γ_x, γ_y は正則化パラメータである. 最大化問題はラグランジュの未定乗数法によって解くことができる. (**??**) 式を $\lambda_{\alpha 2}$, $\lambda_{\beta 2}$ をラグランジュ乗数としたラグランジュ関数は次式で与えられる.

$$\begin{aligned} L_{k2}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda_{\alpha 2}, \lambda_{\beta 2}) = \boldsymbol{\alpha}^T K_x K_y \boldsymbol{\beta} &- \frac{\lambda_{\alpha 2}}{2} (\boldsymbol{\alpha}^T K_x^2 \boldsymbol{\alpha} + \gamma_x \boldsymbol{\alpha}^T \boldsymbol{\alpha} - 1) \\ &- \frac{\lambda_{\beta 2}}{2} (\boldsymbol{\beta}^T K_y^2 \boldsymbol{\beta} + \gamma_y \boldsymbol{\beta}^T \boldsymbol{\beta} - 1). \end{aligned}$$

上式を $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ に関して偏微分し, 0 となる停留点を計算し, 整理すると次式をえる.

$$\begin{pmatrix} O & (K_x^2 + \gamma_x I_n)^{-1} K_x K_y \\ (K_y^2 + \gamma_y I_n)^{-1} K_y K_x & O \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \lambda_{k2} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}.$$

従って, 固有値問題に帰着することができる.

3 スパースカーネル正準相関分析

2 節でのカーネル正準相関分析では非線形のデータに関して対応することができた．次に, 過適合を防ぐためにスパース性を持ったカーネル正準相関分析について述べる．

カーネル行列のランクを $\hat{r} = \mathrm{rank}(K_x)$, $\hat{s} = \mathrm{rank}(K_y)$, $\hat{m} = \mathrm{rank}(K_xK_y)$ とし, K_x, K_y の特異値分解を次のように与える．

$$K_x = \tilde{U}_1\Pi_1\tilde{U}_1^T, \quad K_y = \tilde{V}_1\Pi_2\tilde{V}_1^T.$$

ここで, $\tilde{U} \in \mathbb{R}^{n\times n}$, $\tilde{U}_1 \in \mathbb{R}^{n\times\hat{r}}$, $\tilde{U}_2 \in \mathbb{R}^{n\times(n-\hat{r})}$, $\Pi_1 \in \mathbb{R}^{\hat{r}\times\hat{r}}$, $\tilde{V} \in \mathbb{R}^{n\times n}$ $\tilde{V}_1 \in \mathbb{R}^{n\times\hat{s}}$, $\tilde{V}_2 \in \mathbb{R}^{n\times(n-\hat{s})}$, $\Pi_2 \in \mathbb{R}^{\hat{s}\times\hat{s}}$, とし $\tilde{U}$, $\tilde{V}$ は直交行列であり, Π_1, Π_2 は対角行列である．

次に $\tilde{U}_1^T\tilde{V}_1$ の特異値分解を次式で与える．$\tilde{U}_1^T\tilde{V}_1 = \tilde{P}_1\Pi\tilde{P}_2^T$．ここで, $\tilde{P}_1 \in \mathbb{R}^{\hat{r}\times\hat{r}}$, $\tilde{P}_2 \in \mathbb{R}^{\hat{s}\times\hat{s}}$ は直交行列であり, $\Pi \in \mathbb{R}^{\hat{r}\times\hat{s}}$ は対角行列である．このとき,

$$\boldsymbol{\alpha} = \tilde{U}_1\Pi_1^{-1}\tilde{\boldsymbol{p}}_{1_1} + \tilde{U}_2\mathcal{E}, \quad \boldsymbol{\beta} = \tilde{V}_1\Pi_2^{-1}\tilde{\boldsymbol{p}}_{2_1} + \tilde{V}_2\mathcal{F}$$

とすると, 上式が (**??**) 式の解であることが Chu *et al.* (2013) より知られている．ここで $\mathcal{E} \in \mathbb{R}^{(n-\hat{r})}$, $\mathcal{F} \in \mathbb{R}^{(n-\hat{s})}$ は任意のベクトルである．(**??**) 式が最小 2 乗法と関連があることが Chu *et al.*, (2013) より示される．$\tilde{\boldsymbol{t}}_x$, $\tilde{\boldsymbol{t}}_y$ を次式のように与える．

$$\tilde{\boldsymbol{t}}_x = \tilde{U}_1\tilde{\boldsymbol{p}}_{1_1}, \quad \tilde{\boldsymbol{t}}_y = \tilde{V}_1\tilde{\boldsymbol{p}}_{2_1} \tag{3.1}$$

ただし, $\tilde{\boldsymbol{p}}_{1_1}$ は行列 $\tilde{P}_1$ の 1 列目のベクトルである．$\tilde{\boldsymbol{p}}_{2_1}$ も同様である．(**??**) 式より, $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$ は次式を満たす．

$$\hat{\alpha} = \arg\min\{||K_x\boldsymbol{\alpha} - \tilde{\boldsymbol{t}}_x||^2\}, \quad \hat{\boldsymbol{\beta}} = \arg\min\{||K_y\boldsymbol{\beta} - \tilde{\boldsymbol{t}}_y||^2\}. \tag{3.2}$$

これは (**??**) 式の解である．この (**??**) 式に L_1 ノルムを加えることによりスパース性を持たせることができる．以下の最小 2 乗問題を解くことを考える．

$$\arg\min_{\boldsymbol{\alpha}}\frac{1}{2}||K_x\boldsymbol{\alpha} - \tilde{\boldsymbol{t}}_x||^2 + \lambda_x||\boldsymbol{\alpha}||_1, \quad \arg\min_{\boldsymbol{\beta}}\frac{1}{2}||K_y\boldsymbol{\beta} - \tilde{\boldsymbol{t}}_y||^2 + \lambda_y||\boldsymbol{\beta}||_1.$$

ここで λ_x, $\lambda_y > 0$ は正則化パラメータである．このようにしてカーネル正準相関分析に対してスパース性を持たせることができる．

4 Penalized Matrix Decomposition スパース正準相関分析

線形の正準相関分析ではゲノムデータといった高次元データに対して直接適用させせることは難しい．そのために多くの制約を用いた Wittel *et al.* (2009) による正準相関分析が提案されている．本節ではスパース性を持つ正準相関分析について述べる．

4.1 Penalized Matrix Decomposition

Witten *et al.* (2009) によって提案された Penalized Matrix Decomposition (以後 PMD) について述べる．PMD を用いることにより Tibshirani (1996) によって提案された L_1 ノルムを用いた正則化特異値分解をえることができる．

いま 行列 $X \in \mathbb{R}^{n\times p}$ が与えられたとする．このとき, 行列 X の階数が r のとき行列 X の特異値分解は次式で与えられる．

$$X = UDV^T.$$

ここで, $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r)$, $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r)$, $D = \mathrm{diag}(d_1, \ldots, d_r)$ であり, 行列 D は行列 U, V は正規直交行列である．このとき, 行列 X のランク 1 の近似を考える．すなわち, 次の最適化問題を解く．

$$\arg\min_{d,\,\boldsymbol{u},\,\boldsymbol{v}}\frac{1}{2}||X - d\boldsymbol{u}\boldsymbol{v}^T||_F^2, \quad \text{subject to } ||\boldsymbol{u}||_2^2 = 1,\ ||\boldsymbol{v}||_2^2 = 1,\ ||\boldsymbol{u}||_1 \le z_1,\ ||\boldsymbol{v}||_1 \le z_2. \tag{4.1}$$

Witten*et al.* (2009) によって, (**??**) 式の最適化問題は次式に変形できる.

$$\mathop{\arg\max}_{\boldsymbol{u},\ \boldsymbol{v}} \boldsymbol{u}^T X \boldsymbol{v}, \quad \text{subject to } ||\boldsymbol{u}||_2^2 \leq 1,\ ||\boldsymbol{v}||_2^2 \leq 1,\ ||\boldsymbol{u}||_1 \leq z_1,\ ||\boldsymbol{v}||_1 \leq z_2. \tag{4.2}$$

4.2 PMD によるスパース正準相関分析

このときの相関係数の最大化問題は 4.1 節で述べた PMD を用いる. (**??**) 式において行列 X を $X^T Y$ に置き換えることによって正準相関分析に PMD を適用できる. すなわち次式の最適化問題を解くことになる.

$$\mathop{\arg\max}_{\boldsymbol{a},\boldsymbol{b}} \boldsymbol{a}^T X^T Y \boldsymbol{b}, \quad \text{subject to } \boldsymbol{a}^T\boldsymbol{a} \leq 1,\ \boldsymbol{b}^T\boldsymbol{b} \leq 1,\ ||\boldsymbol{a}||_1 \leq z_1,\ ||\boldsymbol{b}||_1 \leq z_2.$$

4.3 カーネル正準相関分析へ拡張

PMD を適用したスパース正準相関分析をカーネル正準相関分析に拡張する. (**??**) 式を, PMD による正準相関分析に適用させる.

$$\mathop{\arg\max}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{\alpha}^T K_x K_y \boldsymbol{\beta}, \quad \text{subject to } \boldsymbol{\alpha}^T\boldsymbol{\alpha} \leq 1,\ \boldsymbol{\beta}^T\boldsymbol{\beta} \leq 1,\ ||\boldsymbol{\alpha}||_1 \leq z_1,\ ||\boldsymbol{\beta}||_1 \leq z_2.$$

5 まとめと今後の研究課題

本論文では, はじめに線形性に基づく正準相関分析, カーネル関数を用いた非線形正準相関分析の定式化を行った. 次に, L_1 ノルム正則化法による正準相関分析について述べ, 正準相関分析と最小 2 乗法の関係性について述べた. さらにモデルの非線形化とスパース性を融合したスパース非線形正準相関分析を理論的に定式化した.

今後の研究課題として, カーネル関数の選択法, 選択したカーネル関数のパラメータの決定, 正則化パラメータの決定や推定アルゴリズムの開発, 様々な正則化項を用いた正準相関分析の定式化などが今後の研究として挙げられる. また, 正準変量をいくつまで求めるのが適切か, その基準をどのように設定するかなどの問題も検討したい.

参考文献

[1] 赤穂 昭太郎. (2000). カーネル正準相関分析. 2000 年情報論的学習理論ワークショップ (July 2000)

[2] Chu, D., Liao, L. Z., Ng, M. K., and Zhang, X. (2013). Sparse canonical correlation analysis: New formulation and algorithm. *IEEE transactions on pattern analysis and machine intelligence*, **35(12)**, 3050-3065.

[3] Chu, D., Liao, L., Ng, M., and Zhang, X. (2013). Sparse kernel canonical correlation analysis. *In Proceedings of International Multiconference of Engineers and Computer Scientists, Hong Kong.*

[4] Hardoon, D. R., Szedmak, S., and Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, **16(12)**, 2639-2664.

[5] Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28(3/4), 321-377.

[6] Malte, Kuss., and Thore, Graepel. (2003). The geometry of kernel canonical correlation analysis, *MPI-Technical Reports*, URL http://www.kyb.mpg.de/publication.html.

[7] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

[8] Witten, D. M., and Tibshirani, R. J. (2009). Extensions of sparse canonical correlation analysis with applications to genomic data, *Statistical applications in genetics and molecular biology*, **8(1)**, 1-27.

[9] Witten, D. M., Tibshirani, R., and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, kxp008.