

# 混合確率的偏正準相関分析

## Mixture of Partial Canonical Correlation Analysis

河野圭祐 横山裕樹 森裕紀 浅田稔

大阪大学大学院工学研究科 知能・機能創成工学専攻

The Japanese Society for Artificial Intelligence

In this article, we propose Mixture of Probabilistic Partial Canonical Correlation Analysis (MPPCCA) to analyze data that include multiple causal relationships. Partial Canonical Correlation Analysis (PCCA) is a method for multivariate statistics. PCCA evaluates a correlation between two multivariate while considering third multivariate, and has been used as a calculation method of Granger Causality which is one of the causal indexes.

However, PCCA is not applicable to analyze data which contain some different partial canonical correlations. In this paper, we addressed the problem by proposing a Mixture of Probabilistic Partial Canonical Correlation (MPPCCA) and deriving a parameters estimation method by EM algorithm.

We show an experiment with artificial data which indicates our method calculate more accurately than existing method. An actual data is also analyzed by the proposed method. The data was divided by into partial correlation.

### 1. はじめに

偏正準相関分析 [1] は、ある多変量の影響を除いた上で、2つの多変量の相関が大きくなる低次元空間への射影を求める多変量解析手法であり、偏相関分析の多変量への拡張といえる。偏正準相関分析は、時系列データの因果解析などに応用されている [2]。また、高次元少サンプルなデータに対しても安定して計算を行うことを目的とした、確率的な解釈も与えられている [3]。

一方で、データ中の偏正準相関関係は全て一定であるとは限らず、複数の関係が混在している可能性がある。例えば、時系列データの因果解析では、多変量間の因果関係が時間とともに切り替わる状況 (図 1) が考えられる。偏正準相関分析では、このようなデータを正しく解析することはできない。

そこで、本研究では確率的偏正準相関分析 [3] の混合モデルを提案し、EM アルゴリズム [4] による、モデルパラメータの更新式を導出した。提案手法では、複数の偏正準相関関係が混在しているデータについて、同一の偏正準相関関係であることを基準としたデータ分割をすることができる。提案手法の検証として、人工的に発生させた、複数の偏正準相関関係が混在したデータを用いた実験によって、提案手法と既存手法 [5] を比較し、提案手法が既存手法と比べて、データを発生させるときに用いた偏正準相関関係を正しく推定できることを確認した。

### 2. 関連研究

本章では本研究に関連する多変量解析手法について説明する。さらに、偏正準相関分析とグレンジャー因果 [6] との関係についても簡単に述べる。

#### 2.1 正準相関分析

正準相関分析 [7, 8] は二つの多変量の相関関係を分析する際に用いられる手法であり、脳波の解析や天気予報など幅広く応用されている。正準相関分析では二つの多変量の相関が最大になる低次元空間を求め、それぞれの多変量をその空間へ射影し、相関係数を計算することで、二つの多変量間の相関関係を定量化する。二つの多変量  $\mathbf{y}^{(1)} = (y_1^{(1)}, y_2^{(1)}, \dots, y_{m_1}^{(1)})^T \in \mathbb{R}^{m_1}$ ,

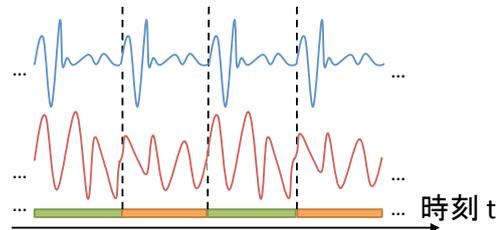


図 1: パターンごとに異なる関係をもつ時系列

$\mathbf{y}^{(2)} = (y_1^{(2)}, y_2^{(2)}, \dots, y_{m_2}^{(2)})^T \in \mathbb{R}^{m_2}$  の間の正準相関分析は以下の一般化固有値問題に帰着する [9]。

$$\begin{aligned} (\Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12} - \rho^2 \Sigma_{22}) \mathbf{u}^{(2)} &= \mathbf{0} \\ (\Sigma_{21}^T \Sigma_{22}^{-1} \Sigma_{21} - \rho^2 \Sigma_{11}) \mathbf{u}^{(1)} &= \mathbf{0} \end{aligned}$$

ここで  $\Sigma_{ab} \in \mathbb{R}^{d_a \times d_b}$  はサンプル  $\mathbf{y}^{(a)} \in \mathbb{R}^{d_a}$ ,  $\mathbf{y}^{(b)} \in \mathbb{R}^{d_b}$  の共分散行列である。また、 $\rho$  は正準相関係数、つまり相関関係の強さである。各固有ベクトルが低次元空間への射影と対応しており、対応する固有値が射影した空間での相関、つまり正準相関に対応する。

#### 2.2 偏正準相関分析

偏正準相関分析 [1] はある一つの多変数の影響を除いた上で二つの多変量の正準相関を求める手法であり、条件を統制した上での正準相関分析や後述するグレンジャー因果 [6] の計算 [2] などに用いられている。多変量  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$  の影響を除いた上で、二つの多変量  $\mathbf{y}^{(1)} = (y_1^{(1)}, y_2^{(1)}, \dots, y_{m_1}^{(1)})^T$ ,  $\mathbf{y}^{(2)} = (y_1^{(2)}, y_2^{(2)}, \dots, y_{m_2}^{(2)})^T$  の間の偏正準相関分析は以下の一般化固有値問題に帰着する [9]。

$$\begin{aligned} (\Sigma_{12|x}^T \Sigma_{11|x}^{-1} \Sigma_{12|x} - \rho^2 \Sigma_{22|x}) \mathbf{u}^{(2)} &= \mathbf{0} \\ (\Sigma_{21|x}^T \Sigma_{22|x}^{-1} \Sigma_{21|x} - \rho^2 \Sigma_{11|x}) \mathbf{u}^{(1)} &= \mathbf{0} \end{aligned} \quad (1)$$

ここで、 $\Sigma_{ab|x} = \Sigma_{ax} - \Sigma_{ax} \Sigma_{xx}^{-1} \Sigma_{xb}$  である。また  $\rho$  は偏正準相関係数であり、 $\mathbf{x}$  の影響を除いた上での、二つの多変量  $\mathbf{y}^{(1)}$ ,  $\mathbf{y}^{(2)}$  の相関関係の強さを表す。

連絡先: 河野圭祐, keisuke.kawano@ams.eng.osaka-u.ac.jp

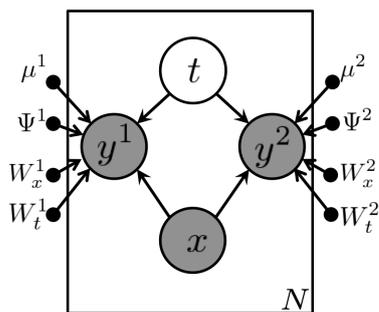


図 2: 確率的偏正準相関分析のグラフィカルモデル [3]

### 2.3 因果指標と偏正準相関分析

偏正準相関分析を用いることで、因果指標の一つであるグレンジャー因果 [6] を求めることができる [2, 10]。グレンジャー因果は時系列データ間の因果関係を示す量であり、経済学や脳神経科学などの分野におけるデータ解析の手法として広く用いられている。グレンジャー因果は時系列データの予測精度に基づく、統計的な因果指標であり、実際の因果関係とは必ずしも一致しない。

二つの多次元時系列データ  $X, Y$  が定常かつ平均 0 のとき、 $Y$  から  $X$  のグレンジャー因果は、以下の一般化固有値問題を解くことで求まる [9]。

$$\left( \begin{array}{c} \Sigma_{X_t Y_{t-k}^{(m)} | X_{t-k}^{(m)}} \Sigma_{X_t X_{t-k}^{(m)} | X_{t-k}^{(m)}}^{-1} \Sigma_{X_t Y_{t-k}^{(m)} | X_{t-k}^{(m)}} \\ -\rho^2 \Sigma_{Y_{t-k}^{(m)} Y_{t-k}^{(m)} | X_{t-k}^{(m)}} \end{array} \right) \mathbf{a} = \mathbf{0} \quad (2)$$

ここで、 $\Sigma_{ab|c} = \Sigma_{ac} - \Sigma_{ac} \Sigma_{cc}^{-1} \Sigma_{cb}$  である。また、 $X_{t-1}^{(m)} = (X_{t-1}^T X_{t-2}^T \dots X_{t-m}^T)^T$ ,  $Y_{t-1}^{(m)} = (Y_{t-1}^T Y_{t-2}^T \dots Y_{t-m}^T)^T$  は埋め込みベクトルである。ここで、式 (2) は、偏正準相関関係を求めたい二つの変数を  $X_t, Y_{t-1}^{(m)}$  とし、影響を除きたい変数を  $X_{t-1}^{(m)}$  とした場合の偏正準相関分析に等しい。

このときグレンジャー因果はそれぞれの固有値  $\rho_i$  を用いて以下のように求まる。

$$G_{Y \rightarrow X} = \frac{1}{2} \log_2 \frac{1}{1 - \rho_i^2} \quad (3)$$

大きい固有値に対応したグレンジャー因果の値は大きくなる。本研究ではこれらのグレンジャー因果のうち、最大のものをグレンジャー因果の値として、解析を行う。

### 2.4 確率的偏正準相関分析

Mukuta et al. [3] は、図 2 に示すグラフィカルモデルにより偏正準相関分析に確率的な解釈を与えた。偏正準相関を考えたい多変量を  $y^{(1)} \in \mathbb{R}^{d_1}$ ,  $y^{(2)} \in \mathbb{R}^{d_2}$ , 影響を除きたい多変量を  $x \in \mathbb{R}^{d_x}$  とするとき、各確率変数の確率密度関数は以下のようになる。

$$\begin{aligned} p(\mathbf{t}_n) &= \mathcal{N}(\mathbf{0}, I_{d_t}) \\ p(\mathbf{y}_n^{(1)} | \mathbf{t}_n) &= \mathcal{N}(W_x^{(1)} \mathbf{x}_n + W_t^{(1)} \mathbf{t}_n + \boldsymbol{\mu}^{(1)}, \Psi^{(1)}) \\ p(\mathbf{y}_n^{(2)} | \mathbf{t}_n) &= \mathcal{N}(W_x^{(2)} \mathbf{x}_n + W_t^{(2)} \mathbf{t}_n + \boldsymbol{\mu}^{(2)}, \Psi^{(2)}) \end{aligned}$$

観測変数  $y^{(1)}, y^{(2)}$  は共通因子  $t$  および、 $x$  から影響を受けており、 $W_t, W_x$  はそれらの強さを表す。偏正準相関分析で

求めた  $d_t$  個の固有値を対角成分にもつ対角行列を  $P_d$ , 対応する固有ベクトルを並べた行列を  $U_{1d}, U_{2d}$  としたとき、パラメータの最尤解は以下ようになる。

$$\begin{aligned} W_x^{(m)} &= \Sigma_{mx} \Sigma_{xx}^{-1} \\ W_t^{(m)} &= \Sigma_{mm|x} U_{dt}^{(m)} M_m \\ \Psi^{(m)} &= \Sigma_{mm|x} - W_t^{(m)} W_t^{(m)T} \\ \boldsymbol{\mu}^{(m)} &= \bar{\mathbf{y}}^{(m)} - W_x^{(m)} \bar{\mathbf{x}} \end{aligned}$$

ただし、 $m = 1, 2$  であり、 $\Sigma_{mm|x} = \Sigma_{mm} - \Sigma_{mx} \Sigma_{xx}^{-1} \Sigma_{xm}$  である。また  $\bar{\mathbf{y}}^{(m)} = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n^{(m)}$ ,  $\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$  を表す。

### 2.5 空間分割による因果パターン抽出

村ら [5] は偏正準相関分析を用いた因果性評価手法において、因果性の原因側の過去情報と結果側の過去、現在情報で構成される特徴空間を想定し、その空間におけるデータ点をクラスタリングすることで、因果性の高いパターンを抽出した。村ら [5] の手法では、埋め込みベクトル  $X_{t-1}^{(m)}, Y_{t-1}^{(m)}$  を用いて特徴点  $\Psi_t = (X_t, X_{t-1}^{(m)}, Y_{t-1}^{(m)})$  が構成される。この特徴点  $\Psi$  には因果性の原因側の過去情報と結果側の過去、現在情報が含まれている。村ら [5] はこれらの特徴点の集合で構成される空間において、ユークリッド距離が近い点同士は同じ因果関係で説明できると仮定し、この空間上で K-means クラスタリングを行い、そのクラスタごとにグレンジャー因果を計算することによって因果性を評価した。しかし、この手法はユークリッド距離が近いデータ点同士は同じ因果関係で説明できると仮定しているため、特徴量は似ているが全く異なる因果関係を持っているデータに対して正しく因果性を評価することができない。

### 3. 確率的偏正準相関分析の混合モデル

本章では、確率的偏正準相関分析 [3] の混合モデルを提案し、EM アルゴリズムによるパラメータおよび潜在変数の最尤推定を行う。提案手法では、同一の偏正準相関関係または同一のグレンジャー因果で説明可能な部分データ集合を抽出し、そのデータ集合ごとに偏正準相関の評価をする。

#### 3.1 生成モデル

偏正準相関関係を求めたい二つの多変量を  $\mathbf{y}^{(1)} \in \mathbb{R}^{m_1}$ ,  $\mathbf{y}^{(2)} \in \mathbb{R}^{m_2}$  とし、影響を取り除きたい多変量を  $\mathbf{x} \in \mathbb{R}^{d_x}$  とする。また、潜在変数として  $\mathbf{t}_n \in \mathbb{R}^{d_t}$  および  $z_{nk} \in \{0, 1\}$  を導入する。ただし  $\min(d_1, d_2) \geq d_t$  である。 $\mathbf{t}_n$  は  $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}$  間の共通因子を表す。また、 $z_{nk}$  は  $n$  番目のデータ点がどのクラスタから生成されたものかを表しており、観測データから  $z_{nk}$  を求めることで、各点がどのクラスタから生成されたものかを推定することができる。 $\sum_k z_{nk} = 1$  であり、 $z_k$  はどれか一つの  $k$  に対して  $z_k = 1$  で、それ以外の  $k$  に対して  $z_k = 0$  である。

$$\begin{aligned} p(\mathbf{t}_n) &= \mathcal{N}(\mathbf{t}_n | \mathbf{0}, I_{d_t}) \\ p(\mathbf{y}_n^{(1)} | \mathbf{t}_n, \mathbf{x}_n; W_{xk}^{(1)}, W_{tk}^{(1)}, \boldsymbol{\mu}_k^{(1)}, \Psi_k^{(1)}, z_{nk}) \\ &= \mathcal{N}(\mathbf{y}_n^{(1)} | W_{xk}^{(1)} \mathbf{x}_n + W_{tk}^{(1)} \mathbf{t}_n + \boldsymbol{\mu}_k^{(1)}, \Psi_k^{(1)})^{z_{nk}} \\ p(\mathbf{y}_n^{(2)} | \mathbf{t}_n, \mathbf{x}_n; W_{xk}^{(2)}, W_{tk}^{(2)}, \boldsymbol{\mu}_k^{(2)}, \Psi_k^{(2)}, z_{nk}) \\ &= \mathcal{N}(\mathbf{y}_n^{(2)} | W_{xk}^{(2)} \mathbf{x}_n + W_{tk}^{(2)} \mathbf{t}_n + \boldsymbol{\mu}_k^{(2)}, \Psi_k^{(2)})^{z_{nk}} \\ p(z_{nk}; \boldsymbol{\pi}_k) &= \text{Multinomial}(z_{nk} | \boldsymbol{\pi}_k) \end{aligned}$$

ここで  $\mathcal{N}$  は、以下のような正規分布の確率密度関数である。

$$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \Psi) = \frac{1}{\sqrt{|2\pi\Psi|}} \exp \left[ -\frac{1}{2} \{(\mathbf{x} - \boldsymbol{\mu})^T \Psi^{-1} (\mathbf{x} - \boldsymbol{\mu})\} \right]$$

ゆえに、提案モデルにおいて、観測変数  $\mathbf{y}_n^{(1)}$  および  $\mathbf{y}_n^{(2)}$  は、影響を取り除きたい変数  $\mathbf{x}_n$  および潜在変数  $t_n$  によって、平均が定まる正規分布に従う変数であるといえる。また、 $\pi_k$  は混合係数で  $0 \leq \pi_k \leq 1$ ,  $p(z_{nk}; \pi_k) = \pi_k$ ,  $\sum_{k=1}^K \pi_k = 1$  である。

### 3.2 EM アルゴリズムによる最尤推定

$\{\mathbf{y}, z\}$  に対する対数尤度関数は以下で表される。

$$\ln p(\mathbf{y}, z \mid \Theta) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \ln \pi_k \mathcal{N}(\mathbf{y}_n \mid W_{xk} \mathbf{x}_n + \boldsymbol{\mu}_k, C_k)$$

ここで  $\Theta$  はパラメータの組であり、 $\Theta = [\pi_k, \boldsymbol{\mu}_k, \Psi_k W_{xk}, W_{tk}]$  である。また、負担率  $r_{nk}$  を以下のように定義する。

$$r_{nk} = \frac{\pi_k \mathcal{N}(\mathbf{y}_n \mid W_{xk} \mathbf{x}_n + \boldsymbol{\mu}_k, C_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{y}_n \mid W_{xj} \mathbf{x}_n + \boldsymbol{\mu}_j, C_j)}$$

EM アルゴリズムによって、パラメータの最尤解を求める。E ステップでは現在のパラメータ  $\Theta^{cur}$  を用いて負担率を求め、完全データに対する対数尤度の期待値  $Q(\Theta, \Theta^{cur})$  を計算する。

$$Q(\Theta, \Theta^{cur}) = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \{ \ln \pi_k + \mathcal{N}(\mathbf{y}_n \mid W_{xk} \mathbf{x}_n + \boldsymbol{\mu}_k, C_k) \}$$

M ステップでは求めた対数尤度の期待値を最大化するパラメータ  $\Theta_{new}$  を求める。

$$\Theta_{new} = \underset{\Theta}{\operatorname{argmax}} Q(\Theta, \Theta^{cur})$$

以上より、パラメータの更新式は以下ようになる。

$$\begin{aligned} r_{nk} &= \frac{\pi_k \mathcal{N}(\mathbf{y}_n \mid W_{xk} \mathbf{x}_n + \boldsymbol{\mu}_k, C_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{y}_n \mid W_{xj} \mathbf{x}_n + \boldsymbol{\mu}_j, C_j)} \\ \boldsymbol{\mu}_k &= \frac{\sum_{n=1}^N r_{nk} (\mathbf{y}_n - W_{xk} \mathbf{x}_n)}{\sum_{n=1}^N r_{nk}} \\ \pi_k &= \frac{1}{N} \sum_{n=1}^N r_{nk} \\ W_{xk} &= \left( \sum_{n=1}^N r_{nk} \tilde{\mathbf{y}}_{nk} \tilde{\mathbf{x}}_{nk}^T \right) \left( \sum_{n=1}^N r_{nk} \tilde{\mathbf{x}}_{nk} \tilde{\mathbf{x}}_{nk}^T \right)^{-1} \\ W_{tk} &= U_k (\Lambda_k - \Psi_k)^{\frac{1}{2}} R \\ \Psi_k &= S_k - W_{tk} W_{tk}^T \end{aligned}$$

ただし、 $\tilde{\mathbf{y}}_{nk} = \mathbf{y}_n - \sum_{n=1}^N \frac{r_{nk} \mathbf{y}_n}{\sum_{n=1}^N r_{nk}}$ ,  $\tilde{\mathbf{x}}_{nk} = \mathbf{x}_n - \sum_{n=1}^N \frac{r_{nk} \mathbf{x}_n}{\sum_{n=1}^N r_{nk}}$ ,  $S_k = \frac{\sum_{n=1}^N r_{nk} (\tilde{\mathbf{y}}_{nk} - W_{xk} \tilde{\mathbf{x}}_{nk}) (\tilde{\mathbf{y}}_{nk} - W_{xk} \tilde{\mathbf{x}}_{nk})^T}{\sum_{n=1}^N r_{nk}}$  であり、 $U_k$  は  $S_k$  の固有ベクトルからなる行列、 $\Lambda_k$  は  $S_k$  の固有値を対角成分に持つ対角行列、 $R$  は任意の直交行列である。

また、提案手法において、2.3 章と同様に、偏正準相関関係を求めたい二つの変数を  $\mathbf{X}_t$ ,  $\mathbf{Y}_{t-1}^{(m)}$  とし、影響を除きたい変数を  $\mathbf{X}_{t-1}^{(m)}$  とすることで、同一のグレンジャー因果で説明可能なデータ集合を取り出すことができる。

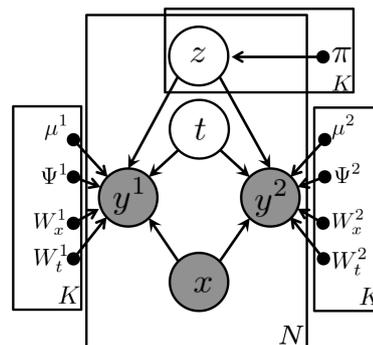


図 3: 混合確率的偏正準相関分析のグラフィカルモデル

## 4. 人工データを用いた評価実験

本節では、提案モデルとパラメータ推定のための EM アルゴリズムが期待通りに求まるかを検討するため、提案モデルに適切なパラメータを代入して人工データを生成し、その人工データからクラスタを推定することを試みる。人工データを生成するために用いる混合クラスタ数  $K = 3$  とする。また、観測データを  $y^{(1)} \in \mathbb{R}^2$ ,  $y^{(2)} \in \mathbb{R}^3$ ,  $x \in \mathbb{R}^3$  とする。今回の実験では、正則化項  $\eta_C$ ,  $\eta_{W_x}$  はともに 0 とした。クラスタ毎のパラメータ  $\mu^{(1)}, \mu^{(2)}, W_x^{(1)}, W_x^{(2)}, W_t^{(1)}, W_t^{(2)}, \Psi^{(1)}, \Psi^{(2)}$ , 混合係数  $\pi$  を表 1 の値に設定し、生成モデルからデータを作成する。

生成したデータを提案手法である混合確率的偏正準相関分析および朴らの手法 [5] によって解析する。生成モデルから 1000 点のデータを作成し、それぞれの手法でクラスタリングを行うまでを 1 試行とする。ここで、あるデータ点  $x_n$  に対する真のクラスタを  $k_n^*$ , 解析によって得られたクラスタを  $k_n$  とする。提案手法と既存手法によって得られた結果を比較するために、各試行における真のクラスタを再現することができなかった割合、クラスタ非再現率を求める。クラスタ非再現率を定義するために、以下のように多数派クラスタ  $\hat{k}(k)$  を定義する。

$$\hat{k}(k) = \underset{j}{\operatorname{argmax}} |C_k \cap C_j^*|$$

ここで、 $C_k^* = \{x_n \mid k_n^* = k\}$  は、ある真のクラスタに所属するデータ集合、 $C_k = \{x_n \mid k_n = k\}$  は、解析によって得られたあるクラスタに所属するデータ集合である。多数派クラスタは、解析によって得られたクラスタに、最も多く存在していた真のクラスタを表している。多数派クラスタを用いて、クラスタ非再現率を以下のように求める。

1. あるデータ点  $x_n$  に対して、多数派クラスタ  $\hat{k}_n$  と真のクラスタ  $k_n^*$  が一致していないとき、このデータ点を誤り点とする。
2. 全てのデータ点に対して誤り点を求め、全てのデータ点に対して誤り点が占める割合をクラスタ非再現率とする。

クラスタ非再現率は、提案手法と真のクラスタの対応関係を多数派クラスタという形で求めた上での、多数派クラスタに当てはまらない点の割合といえる。

それぞれの手法で 1000 試行の解析を行ったときのクラスタ非再現率のヒストグラムを図 4 に示す。図より、朴らの手法 [5] と比較して、提案手法によって正しいクラスタを推定できていることが確認できる。

表 1: 人工データのパラメータ (生成モデルからサンプリングした人工データを用いた実験)

k	$\mu^{(1)}$	$\mu^{(2)}$	$W_x^{(1)}$	$W_x^{(2)}$
1	$\begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}$	$\begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.2 & 0.3 & 0.4 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.2 & 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.5 & 0.6 \end{pmatrix}$
2	$\begin{pmatrix} -0.5 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 1.2 \\ 4.3 \\ 0.4 \end{pmatrix}$	$\begin{pmatrix} 0.5 & 0.4 & -0.3 & 0.4 \\ -0.1 & 0.3 & 0.3 & -0.5 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.2 & -0.3 & 4.5 \\ 5.2 & -0.3 & 0.4 & 1.4 \\ -0.3 & 0.4 & 0.5 & -0.6 \end{pmatrix}$
3	$\begin{pmatrix} 0.1 \\ 4.3 \end{pmatrix}$	$\begin{pmatrix} 2.2 \\ -4.3 \\ 2.4 \end{pmatrix}$	$\begin{pmatrix} -1.1 & 0.7 & -0.8 & 0.4 \\ 0.2 & 2.3 & 0.4 & 0.5 \end{pmatrix}$	$\begin{pmatrix} -0.4 & 0.7 & 0.3 & -0.4 \\ -0.2 & 0.3 & -0.4 & 0.5 \\ 0.3 & 0.1 & 0.1 & 0.9 \end{pmatrix}$

k	$W_t^{(1)}$	$W_t^{(2)}$	$\Psi^{(1)}$	$\Psi^{(2)}$	$\pi$
1	$\begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 0.3 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 0.3 \\ 0.3 & 0.4 \end{pmatrix}$	$\begin{pmatrix} 0.05 & 0.08 \\ 0.08 & 0.13 \end{pmatrix}$	$\begin{pmatrix} 0.21 & 0.24 & 0.32 \\ 0.24 & 0.29 & 0.40 \\ 0.32 & 0.40 & 0.57 \end{pmatrix}$	0.33
2	$\begin{pmatrix} 1.1 & 0.2 \\ 0.2 & 2.3 \end{pmatrix}$	$\begin{pmatrix} 4.1 & 0.2 \\ 3.2 & 0.3 \\ -2.3 & 0.4 \end{pmatrix}$	$\begin{pmatrix} 0.40 & 0.18 \\ 0.18 & 0.13 \end{pmatrix}$	$\begin{pmatrix} 0.89 & 0.44 & 0.80 \\ 0.44 & 0.29 & 0.44 \\ 0.80 & 0.44 & 0.77 \end{pmatrix}$	0.33
3	$\begin{pmatrix} -1.0 & 0.2 \\ 0.2 & -0.5 \end{pmatrix}$	$\begin{pmatrix} 0.0 & -0.6 \\ 0.9 & -0.9 \\ 0.7 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 0.09 \end{pmatrix}$	$\begin{pmatrix} 2.93 & 1.84 & 2.93 \\ 1.84 & 1.69 & 1.96 \\ 2.93 & 1.96 & 6.9 \end{pmatrix}$	0.34

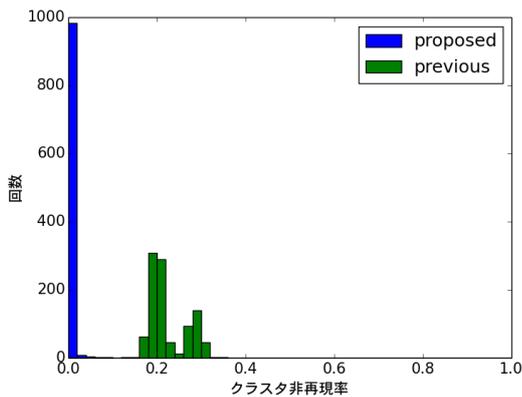


図 4: クラスタ非再現率のヒストグラム

## 5. おわりに

本研究では、複数の偏正準相関関係が混在したデータに対して、それぞれの偏正準相関関係ごとにデータを分割する手法を提案した。また、提案手法を人工データに適用し、既存手法 [5] と比較して、より正しくクラスタを推定できることを確かめた。紙幅の関係で紹介しなかったが、2 者間のコミュニケーションに関する実データ解析も行っている。本研究の発展として、クラスタ数を自動で推定するモデルなどが考えられる。

## 参考文献

[1] B. Raja Rao. Partial canonical correlations. *Trabajos de estadística y de investigación operativa*, Vol. 20, No. 2, pp. 211–219, 1969.

[2] André Fujita, Joao Ricardo Sato, Kaname Kojima, Luciana Rodrigues Gomes, Masao Nagasaki, Mari Cleide Sogayar, and Satoru Miyano. Identification of granger causality between gene sets. *Journal of Bioinformatics*

and Computational Biology, Vol. 8, No. 04, pp. 679–701, 2010.

[3] Yusuke Mukuta and Tatsuya Harada. Probabilistic Partial Canonical Correlation Analysis. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 1449–1457, 2014.

[4] ビショップ CM. パターン認識と機械学習. シュプリンガー・ジャパン, 2008.

[5] 朴鍾範, 森裕紀, 浅田稔. Granger causality に基づく多次元時系列データからの causal pattern 抽出. システム制御情報学会研究発表講演会講演論文集, 2014. Vol.DVD-ROM.

[6] Clive W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[7] Harold Hotelling. Relations between two sets of variates. *Biometrika*, pp. 321–377, 1936.

[8] David Weenink. Canonical correlation analysis. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, Vol. 25, pp. 81–99, 2003.

[9] Yuya Yamashita, Tatsuya Harada, and Yasuo Kuniyoshi. Causal flow. In *IEEE International Conference on Multimedia and Expo*, pp. 1–6. IEEE, 2011.

[10] João R. Sato, André Fujita, Elisson F. Cardoso, Carlos E. Thomaz, Michael J. Brammer, and Edson Amaro Jr. Analyzing the connectivity between regions of interest: an approach based on cluster granger causality for fmri data analysis. *Neuroimage*, Vol. 52, No. 4, pp. 1444–1455, 2010.