

# ジョイントモデリングに基づく多変量関数クラスタリングと 気象データへの応用

横浜市立大学 医学部 三角 俊裕  
滋賀大学 データサイエンス学部 松井 秀俊  
中央大学 理工学部 小西 貞則

## 1. 概要

経時的に測定された多変量データの解析を目的として、非線形混合効果ジョイントモデリングに基づく多変量関数クラスタリング手法を提案する。非線形混合効果ジョイントモデルは、各個体の測定時点が異なるデータを容易に取り扱い、変量間の相関も考慮することができる一方で、多数のパラメータがモデルに含まれるため高い計算コストを要する。この問題を克服するため、pairwise fitting アプローチ (Fieuws and Verbeke, 2006) を利用したモデル推定を行う。本モデル推定により得られる個体ごとのランダム効果関数の予測係数ベクトルに対して、自己組織化マップを適用することでクラスタリングを行う (松井他, 2016)。本提案手法は、2000 年から 2017 年の間に発生した台風のデータ解析に適用する。

## 2. 非線形混合効果ジョイントモデル

$n$  個の  $K$  変量経時測定データを  $\{y_{ijk}; i = 1, \dots, n, j = 1, \dots, J_i, k = 1, \dots, K\}$  とする。ここで、 $y_{ijk}$  は第  $i$  個体の  $t_{ij}$  時点での第  $k$  変量に関する観測値であり、 $J_i$  は第  $i$  個体の観測時点数を表す。このデータに対して、基底展開法に基づく非線形混合効果モデルは以下で与えられる。

$$y_{ijk} = \sum_{l=1}^{p_f} v_{kl} \varphi_l(t_{ijk}) + \sum_{m=1}^{p_r} w_{ikm} \psi_m(t_{ijk}) + \varepsilon_{ijk} = \boldsymbol{\varphi}(t_{ijk})' \boldsymbol{\alpha}_k + \boldsymbol{\psi}(t_{ijk})' \mathbf{b}_{ik} + \varepsilon_{ijk}. \quad (1)$$

ここで、 $\boldsymbol{\varphi}(t_{ijk}) = (\varphi_1(t_{ijk}), \dots, \varphi_{p_f}(t_{ijk}))'$ 、 $\boldsymbol{\psi}(t_{ijk}) = (\psi_1(t_{ijk}), \dots, \psi_{p_r}(t_{ijk}))'$  はそれぞれ平均効果項、ランダム効果項に対する基底関数からなるベクトル、 $\boldsymbol{\alpha}_k = (v_{k1}, \dots, v_{kp_f})'$ 、 $\mathbf{b}_{ik} = (w_{ik1}, \dots, w_{ikp_r})'$  はそれぞれ平均効果項、ランダム効果項の係数ベクトルとする。このとき、第  $i$  個体の  $K$  変量経時測定データに対する非線形混合効果ジョイントモデルは、混合効果モデル表現を利用して以下で与えられる。

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i. \quad (2)$$

ただし、 $\mathbf{y}_i = (\mathbf{y}'_{i1}, \dots, \mathbf{y}'_{iK})'$ 、 $\mathbf{y}_{ik} = (y_{i1k}, \dots, y_{iJ_i k})'$ 、 $\mathbf{X}_i = \text{blockdiag}\{\mathbf{X}_{i1}, \dots, \mathbf{X}_{iK}\}$ 、 $\mathbf{X}_{ik} = (\boldsymbol{\varphi}(t_{i1k}), \dots, \boldsymbol{\varphi}(t_{iJ_i k}))'$ 、 $\mathbf{Z}_i = \text{blockdiag}\{\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{iK}\}$ 、 $\mathbf{Z}_{ik} = (\boldsymbol{\psi}(t_{i1k}), \dots, \boldsymbol{\psi}(t_{iJ_i k}))'$ 、 $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_K)'$ 、 $\mathbf{b}_i = (\mathbf{b}'_{i1}, \dots, \mathbf{b}'_{iK})'$ 、 $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}'_{i1}, \dots, \boldsymbol{\varepsilon}'_{iK})'$ 、 $\boldsymbol{\varepsilon}_{ik} = (\varepsilon_{ijk}, \dots, \varepsilon_{iJ_i k})'$  とする。 $\mathbf{b}_i$  は、平均ベクトル  $\mathbf{0}$ 、分散共分散行列  $\boldsymbol{\Delta}$  の  $Kp_r$  変量正規分布に従う確率変数ベクトルとする。ここで、 $\boldsymbol{\Delta}$  は無構造を仮定しており、変量間の相関が考慮される。また、 $\boldsymbol{\varepsilon}_i$  は互いに独立に平均ベクトル  $\mathbf{0}$ 、分散共分散行列  $\boldsymbol{\Sigma} \otimes \mathbf{I}_{J_i}$  の  $KJ_i$  変量正規分布に従う観測誤差とする。ここで、 $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_K^2)$ 、 $\mathbf{I}_{J_i}$  は  $J_i$  次単位行列である。

## 3. Pairwise fitting アプローチに基づくモデル推定

非線形混合効果ジョイントモデル (2) に含まれる未知パラメータは  $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \sigma_1^2, \dots, \sigma_K^2, (\text{vech} \boldsymbol{\Delta})')'$  であり、その数は、変量数 ( $K$ ) および基底関数の個数 ( $p_f, p_r$ ) に依存し、パラメータ数が合計で  $\kappa = (p_f + 1)K + p_r(p_r + 1)K(K + 1)/4$  の大規模モデルとなる。このため、各 2 変量非線形混合効果モデルに対して EM アルゴリズムを用いて推定を行った後、すべての推定量の組を統合してフルモデルを推定する。モデリングの詳細と台風データの適用例は、当日報告する。

## 参考文献

- Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* **62**:424–431  
松井秀俊, 三角俊裕, 横溝孝明, 小西貞則. (2016). 非線形混合効果モデルに基づく関数データクラスタリング. 応用統計学 **44**, 25–45.