

偏った時系列データを用いた機械学習による時間発展モデリング

中井 拳吾¹, 齊木 吉隆²

¹ 岡山大学 学術研究院 環境生命自然科学学域, ² 一橋大学大学院 経営管理研究科経営管理専攻
e-mail: knakai@okayama-u.ac.jp

1 概要

リザーバーコンピューティングと呼ばれる機械学習手法が決定論的ダイナミクスの時系列モデリングに有効であり [1]、時系列予測や力学系構造の再現性 [2, 3] についても明らかになってきている。現実の現象のデータを学習する場合を考えると、準備できるデータが少ない場合やデータに偏りがある場合が多くある。本講演では周期軌道を用いて偏った学習データを生成し、構成される機械学習モデルへの影響についてアトラクターや不変密度分布等の再現性に注目して明らかにする [4]。

2 リザーバーコンピューティングの概要

力学系 $d\phi/dt = \mathbf{f}(\phi)$ の変数 $\mathbf{u} = \mathbf{h}_1(\phi) \in \mathbb{R}^M$ について考える。この変数はある時刻までは既知とし、入力 $\mathbf{u}(t)$ に対して出力が $\mathbf{u}(t + \Delta t)$ となるように $\mathbf{u}(t)$ と $\mathbf{u}(t + \Delta t)$ の対応関係を決定する。 Δt は時間発展の時間ステップである。リザーバーコンピューティングではニューラルネットワークを用いるが、従来の機械学習とは異なりニューラルネットワークそのものは学習しない。このため学習のために要する計算量が極めて少ない事が特徴である。これによりニューラルネットワークの次元を増やすことに注力することが可能になり高い表現力が実現できているため、複雑な挙動を示すダイナミクスであってもそれが決定論的であるならばこの機械学習手法が有効である。

3 結果

ローレンツ方程式 $\dot{x} = -10x + 10y$, $\dot{y} = -xz + 28x - y$, $\dot{z} = xy - 8z/3$ について考える。学習データに偏りのある場合に、得られる機械学習モデルがその偏りの影響をどのように受けるかを明らかにする。考察を行いやすくするために不安定周期軌道を学習データとして用いた。以下で登場する周期は対応するポアンカレ写像の整数値周期を表す。学習の詳細については [4, 5] を参照されたい。

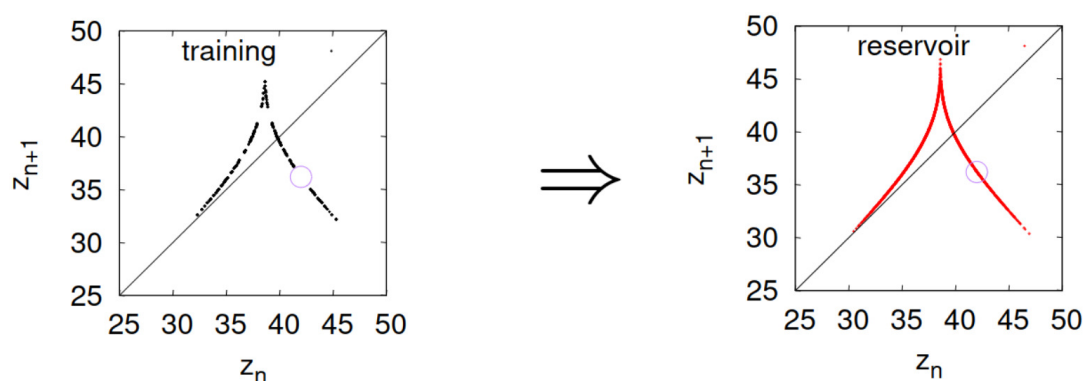


図 1. 周期 9 までの周期軌道のデータセットから中心 $(42, 36.2)$ で半径が 1 の円 $B_1(42, 36.2)$ の内部を通るものを除外して得た周期軌道のデータセット (左) とそれを学習して得た機械学習モデル (右) のリターンプロット (z_n, z_{n+1}) をそれぞれ描いた。ただし、 z_n は z 変数の極大値を意味する。円 $B_1(42, 36.2)$ も書いた。学習データには円の内部の情報は無いが、構成した機械学習モデルはその円の内部の情報を再構成できている (図は [4]Fig.3 引用)。

周期が 9 までの周期軌道のデータセットを準備し、その周期軌道のうち (z_n, z_{n+1}) 上で中心 $(42, 36.2)$ で半径 1 の円 $B_1(42, 36.2)$ の内部を通過する周期軌道を除いた周期軌道リストを作成する。この周期軌道リストをすべて学習し機械学習モデルを構成する。図 1 に学習に用いた周期軌道のリターンプロットと、これを学習して得た機械学習モデルのリターンプロットをそれぞれ描いた。モデルが円 $B_1(42, 36.2)$ 内の情報を再構築していることがわかる。

講演では短い周期の周期軌道のみを学習して正解のカオス的振る舞いを再構築する様子や、一部のデータを過剰に学習した場合の機械学習モデルに与える影響についても言及する。

謝辞 本研究は科研費 (課題番号 :22K17965, 19KK0067, 21K18584, 23H04465, 24K00537) の助成を受けたものである。また、本研究の一部は学際大規模情報基盤共同利用・共同研究拠点 (JHPCN) (課題番号: jh240051)、スーパーコンピュータ共同研究制度 (若手・女性研究者奨励枠) の支援による。

参考文献

- [1] Zhixin Lu, Jaideep Pathak, Brian Hunt, Michelle Girvan, Roger Brockett, and Edward Ott. Reservoir observers: Model-free inference of unmeasured variables in chaotic systems. *Chaos*, 27(4):041102, 2017.
- [2] Miki U. Kobayashi, Kengo Nakai, and Yoshitaka Saiki. Lyapunov analysis of data-driven models of high dimensional dynamics using reservoir computing: Lorenz-96 system and fluid flow. *Journal of Physics: Complexity*, 5(2):025024, 2024.
- [3] Miki U. Kobayashi, Kengo Nakai, Yoshitaka Saiki, and Natsuki Tsutsumi. Dynamical system analysis of a data-driven model constructed by reservoir computing. *Physical Review E*, 104(4):044215, 2021.
- [4] Kengo Nakai and Yoshitaka Saiki. Data-driven modeling from biased small training data using periodic orbits. *arXiv preprint arXiv:2407.06229*, 2024.
- [5] Akane Ookubo and Masanobu Inubushi. Reservoir computing with generalized readout based on generalized synchronization. *arXiv preprint arXiv:2405.14885*, 2024.

Belousov Zhabotinsky 反応を用いたリザーバー計算の数理モデリング

田中 吉太郎¹, 江藤 亘啓¹, 豊田 和人¹, 高木 清二¹, 香取 勇一¹, 櫻沢 繁¹

¹ 公立はこだて未来大学 システム情報科学部

e-mail : y-tanaka@fun.ac.jp

1 はじめに

生物の脳では、神経細胞であるニューロンが互いに複雑に接続し、神経回路網を形成している。この神経回路網における情報処理の様子を数学的に表現したモデルの一つにリザーバー計算がある。リザーバー計算は一般的に、入力層と中間層、出力層から成り、中間層でのニューロンの結合を許す。そして、パラメーターの調整（学習）を中間層から出力層への重みだけに適用するモデルのことをいう。特に、リザーバー計算の中間層を一つの力学系と考えることにより、中間層を物理現象に置き換えた物理リザーバー計算が提案されている。例えば中嶋ら [1] によって、タコ腕に模したソフトマテリアルを用いた物理リザーバー計算、また野津ら [2] によって、円柱後方の流体の流れを用いた物理リザーバー計算、そして香取ら [3] によって、マウスの脳の神経細胞を用いた物理リザーバー計算が提案されている。本研究では、化学反応によって生物の脳で行われているような時系列の学習や予測を再現できるかを調べるために、酸化還元反応が周期的に起こる Belousov Zhabotinsky (BZ) 反応を用いた物理リザーバー計算を提案する。BZ 反応モデルの分岐パラメーターを調整することで、単純な周期的な時系列やカオス的な時系列が有限時刻でおおよそ近似及び予測できることを発表する。

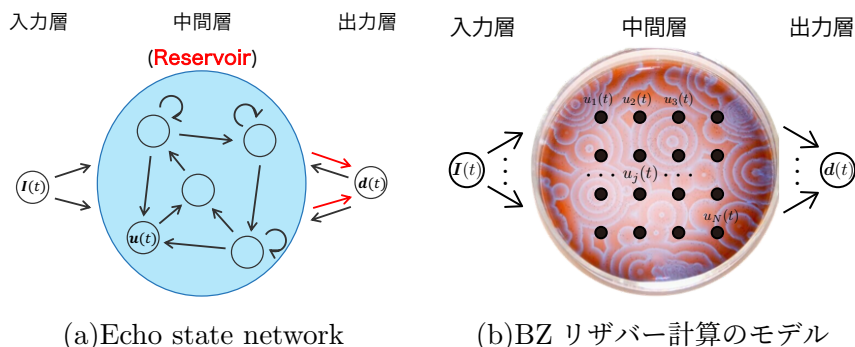


図 1. リザーバー計算と BZ リザーバーの概念図。 $I(t)$ は入力、 $d(t)$ は出力を表す。

2 数理モデリング

この章では、BZ 反応を中間層に置いたリザーバー計算の数理モデルについて説明する。シャーレ $\Omega := (0, L)^2$ 上で進行する BZ 反応を考える。時刻 $t > 0$, 位置 $x \in \Omega$ 上の亜臭素酸の濃度を $u = u(x, t)$, セリウムイオンを $z = z(x, t)$, 臭化マロン酸の濃度を $v = v(x, t)$ とおき、以下の反応拡散系を考える。

$$\begin{cases} \frac{du}{dt} = d_u \Delta u + F(u, z, v), \\ \frac{dz}{dt} = d_z \Delta z + G(u, z, v, \gamma, \{\chi_i\}_i), \\ \frac{dv}{dt} = d_v \Delta v + H(u, z, v). \end{cases} \quad \Omega \times \{t > 0\}, \quad (1)$$

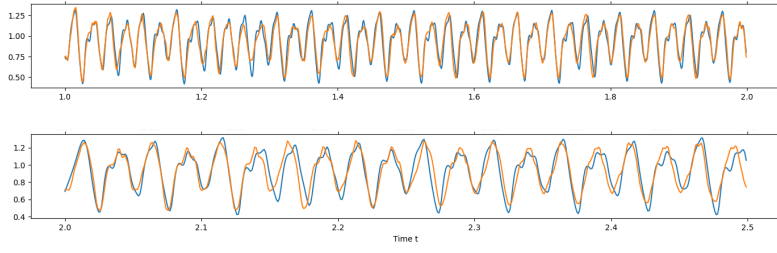


図 2. 学習及び予測タスクの数値計算結果. GF モデルのパラメーター $k_f = 2.633 \times 10^{-3}$. それ以外 [4] に同じ. $d_u = 1, d_z = 0.6, d_v = 1$.

ここで d_u, d_z, d_v は拡散係数であり, 非線形項 F, G, H は Györgyi と Field によって提案された 3 変数系の BZ 反応の数値モデル (GF モデル) によって与える [4]. 境界条件は斉次 Neumann 境界条件を課す. BZ 反応は光照射によって酸化状態が抑制されることが知られており, 学習に用いる教師データの入力や出力の逆投影を光照射で行うことにする. このために, 光が照射される領域を中心を (l_i, m_i) , 半径を $r > 0$ とする円 Ω_i で与え, Ω_i で値をもつ特性関数を χ_i で定義する. 教師データ $\gamma(t)$ は Ω_i で値を持つように非線形項に組み込む. シャーレ Ω 上から格子状に点列 $\{\mathbf{x}\}_{i=1}^N$, ($N \in \mathbb{N}$) をとり, その点列上の u または z の濃度を一つのニューロンに見立てる (図 1(b)). $\mathbf{u}(t) := (u_1, \dots, u_N)^T(t) := (u(\mathbf{x}_1, t), \dots, u(\mathbf{x}_N, t))^T$ (T は転置を表す) とし, 中間層から出力層の重みを $\mathbf{W}^{out} := (W_1^{out}, \dots, W_N^{out})$ とおいて, 出力を $\mathbf{W}^{out} \mathbf{u} = \sum_{i=1}^N W_i^{out} u_i$ で与える. これを用いて, 各時間ステップで教師データと出力が等しくなるように学習を行う:

$$\mathbf{W}^{out} \mathbf{u}(t_j) = \gamma(t_j), \quad j = 0, \dots, \exists M.$$

この連立方程式から \mathbf{W}^{out} を導いた後, 教師データ $\gamma(t_j)$ と出力 $\mathbf{W}^{out} \mathbf{u}(t_j)$ を見比べて, 学習が成功したか近似精度を検証する. 学習が成功した場合は, (1) の γ を $\mathbf{W}^{out} \mathbf{u}$ に置き換えて, 自律系でも教師データに近い時系列が出力できるか, つまり時系列の予測ができるか調べる.

3 数値計算結果

教師データ γ を Mackey-Grass 方程式の解で与え, (1) のリザーバー計算で学習及び予測ができるかを検証した. 図 2 がその数値計算結果である. カオス的な時系列に関して, 有限時刻まである程度の精度で近似及び予測できることがわかった.

4 まとめ

BZ 反応を駆動力とした物理リザーバー計算の数値モデリングを行った. 今後は本モデルの数理的な構造や, 学習が可能となる条件を調査していく予定である.

謝辞 本研究は, 科研費 (課題番号:21H03512) の助成を受けたものである.

参考文献

- [1] Kohei Nakajima (2020). Jpn. J. Appl. Phys. 59 060501, pp. 1-13.
- [2] Ken Goto, Kohei Nakajima, Hirofumi Notsu (2021), New J. Phys. 23 063051
- [3] Sumi et al., Proc. Natl. Acad. Sci. USA., 2023
- [4] L. Györgyi and R. J. Field, 355, NATURE, pp.808-810 (1992).

Littlewood-Paley 分解の数学理論に立脚したリザーバーコンピューティングの実装

神野拓哉¹, 三ツ井孝仁², 中井拳吾³ 齊木吉隆⁴, 米田剛⁴

¹ 富山大学, ² 順天堂大学, ³ 岡山大学, ⁴ 一橋大学

e-mail : t.yoneda@r.hit-u.ac.jp

本講演では、エルニーニョ現象を的確に表す時系列データ Nino 4.3 を使った予測研究結果を紹介する。時系列データに対する学習モデルの将来予測性能を評価するために、ピアソン相関係数 (PCC) が一般的に用いられている。Hassanibesheli-Kurths-Boers[1] では、リザーバーのノード数を 600 と設定し、19 か月先予測で PCC が 0.5 以上となるリザーバー機械学習モデル構築に成功した。

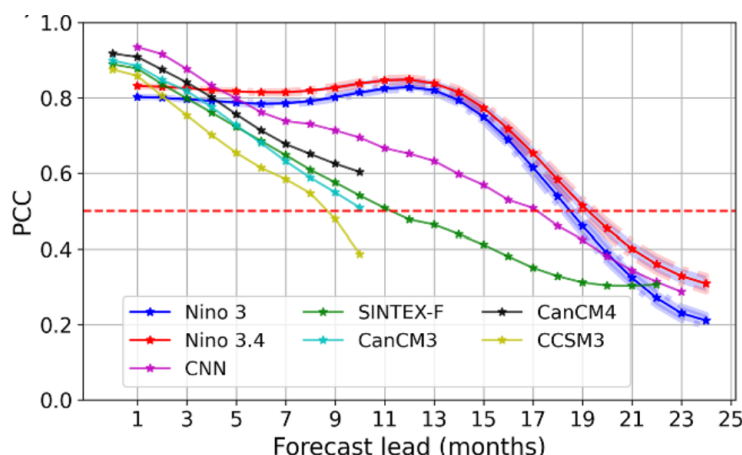


図 1. 先行結果 (赤色のグラフ)

しかし、彼らが用いているフィルタは、将来データも取り込んでしまうものであるため、このモデルが現実のエルニーニョ発生予測に直接使えるわけではない。それを踏まえた上で、我々は将来データを一切取り込まない real time bandpass filter (移動平均の重み関数) を開発した (real time bandpass filter の詳細に関しては [2] を参照されたい) :

$$\Psi_{r_+, r_-, c}(t) = \left(\frac{\sin\left(\frac{t}{r_+}\right)}{t} - \frac{\sin\left(\frac{t}{r_-}\right)}{t} \right) \left(\frac{\sin\left(\frac{t}{c} - \pi\right)}{\frac{t}{c} - \pi} \right) 1_{(-\infty, 0)}(t)$$

通過帯域は $[r_-, r_+]$ であり、 c は重心を調整するパラメタとなる。エルニーニョ現象のスケールが、主に 12 か月から 120 か月までにおさまらるだろう、という気象学的見解を勘案して $(r_+, r_-) = (120/\pi, 12/\pi)$ と設定した (重心が原点に近くなるように $c = 7$ とした)。図 2 を参照されたい。

注意 1 これらフィルタは Littlewood-Paley 分解の数学理論に基づいている。フーリエ変換の不確定性原理を勘案すると、[1] のフィルタは将来データを取り込んでしまうこと、逆に我々のフィルタの通過帯域があいまいなものになってしまうこと、こういった不完全さは避けて通れない。諦念をもって受け入れるしかないのである。

我々の研究では、フィルタをかけたデータをリザーバーコンピューティングによって機械学習させたのち、以下のデコーダを使って元のスケールへと復元し、それを予測データとしている。

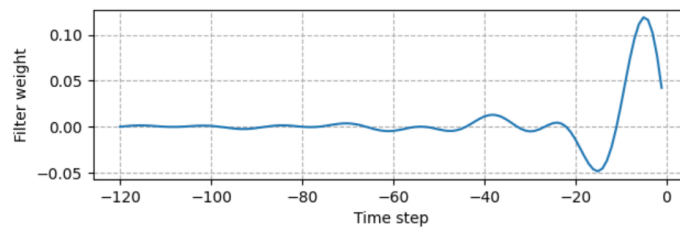
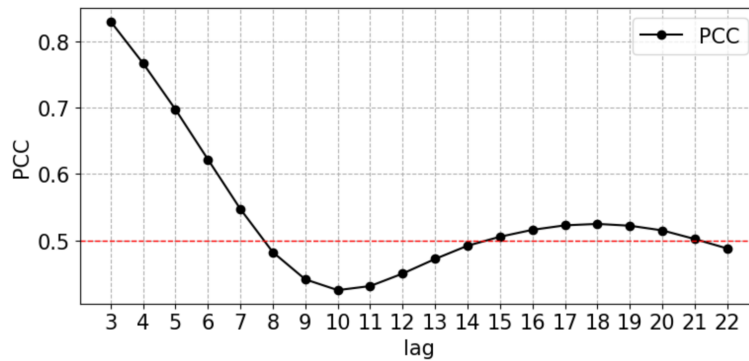


図 2. エルニーニョ現象を（可能な限り）効率的に抽出するための real time bandpass filter

デコーダ: u を時系列データとする。これは離散データなので、次のような逆作用素 Ψ^{-1} が構成できる（詳しくは [3] を参照されたい）。

$$\Psi^{-1}(\Psi * u) = u$$

21 か月先予測に対して PCC が 0.5 以上を実現するようにリザーバー機械学習（およびデコーダを適用）させた結果が下のグラフである。



なお、real time bandpass filter とそれに対応するデコーダの開発以外にも、乱数のシード値に対して次の工夫を凝らした。

- ベイズ最適化に基づく optuna を採用し、ハイパーパラメタも学習させるアルゴリズムを実装した。
- そういったベイズ最適化を適用させるために、乱数のシード値を連続化させなければならなかった。具体的には、自然数上のシード値で生成されるリカレント結合重み行列、入力結合重み行列に対して線形補間を施した（その線形補間によって、シード値の連続性が行列の連続性へと反映される）。

なお、この数値計算結果は、Google Colaboratory という低スペック仮想マシンに基づいており、従って、ベイズ最適化によって得られたリザーバーノード数も 222 と非常に低い。今後、高スペックマシンを積極的に用いることで、より良い学習結果が得られることが期待される。

参考文献

- [1] Hassanibesheli F., Kurths J. and Boers N., *Long-term ENSO prediction with echo-state networks* Environ. Res.: Clim. 1, (2022), 011002.
- [2] Suematsu T., Nakai K., Yoneda T., Takasuka D., Jinno T., Saiki Y., Miura H., *Machine learning prediction of the MJO extends beyond one month*, arXiv:2301.01254
- [3] GitHub, <https://github.com/tsuyoshi-yoneda-math/Autoencoder-in-time>.

因子分析の数理と応用

廣瀬 慧¹

¹ 九州大学マス・フォア・インダストリ研究所

e-mail : [hirose@imi.kyushu-u.ac.jp]

1 概要

因子分析は、観測される多変量の変数から観測できない因子を見出す多変量解析の手法であり、ニューラルネットワークなど潜在変数を含む機械学習の基となるモデルである。このモデルは、もともと最尤法および因子回転によってパラメータが推定されてきたが、高次元データに対してより解釈しやすい因子を推定するためには、正則化法が有用である。本発表では、正則化法と因子回転の関係について述べ、高次元データに対する正則化法の有用性について解説する。

2 因子分析モデル

いま、観測変数ベクトル $\mathbf{X} = (X_1, \dots, X_p)^T$ を平均ベクトル $\mathbf{0}$ 、共分散行列 Σ を持つ p 次元多変量正規分布に従うとする。このとき、因子分析モデルは次のように表される。

$$\mathbf{X} = \Lambda \mathbf{F} + \boldsymbol{\varepsilon}, \quad (1)$$

ここで、 $\Lambda = (\lambda_{ij})$ は $p \times m$ の因子負荷行列、 $\mathbf{F} = (F_1, \dots, F_m)^T$ は共通因子ベクトル、 $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)^T$ は独自因子ベクトルである。ここで、 $E(\mathbf{F}) = \mathbf{0}$ 、 $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ 、 $E(\mathbf{F}\mathbf{F}^T) = \mathbf{I}_m$ 、 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \Psi$ 、 $E(\mathbf{F}\boldsymbol{\varepsilon}^T) = \mathbf{0}$ を仮定する。ただし、 \mathbf{I}_m は $m \times m$ の単位行列とし、 Ψ は $p \times p$ の対角行列とする。 Ψ の対角要素は独自分散と呼ばれる。これらの仮定の下で、観測ベクトル \mathbf{X} の共分散行列は、 $\Sigma = \Lambda\Lambda^T + \Psi$ となる。

3 正則化法と因子回転の関係

n 個の観測値を $\mathbf{x}_1, \dots, \mathbf{x}_n$ とし、対応する標本共分散行列を $\mathbf{S} = (s_{ij})$ とする。パラメータベクトルを $\boldsymbol{\theta} = (\text{vec}(\Lambda)^T, \text{diag}(\Psi)^T)^T$ とする。負の対数尤度関数 $\ell(\boldsymbol{\theta})$ の最大化によって最尤推定量 $\hat{\Lambda}_{ML}$, $\hat{\Psi}_{ML}$ を求めることを考える。

しかし、因子分析モデルには回転の不定性が存在する。いま、任意の直交行列 T に対し、因子負荷行列 Λ と ΛT は、同じ共分散行列 $\Sigma = \Lambda\Lambda^T + \Psi = (\Lambda T)(\Lambda T)^T + \Psi$ を持つ。そこで、適当な損失関数 $Q(\Lambda)$ を用いて、以下の最適化問題を解くことにより、解釈しやすい因子を推定する。

$$\min_{\Lambda, \Psi} Q(\Lambda), \quad \text{subject to } \ell(\Lambda, \Psi) = \hat{\ell} \quad (2)$$

ここで、 $\hat{\ell} = \ell(\hat{\Lambda}_{ML}, \hat{\Psi}_{ML})$ である。この方法は因子回転と対応する。因子回転は、三四半世紀前から使われてきた推定手法である。

ここで、よりスパースな解を得るために次のように問題を修正する。

$$\min_{\Lambda, \Psi} Q(\Lambda), \quad \text{subject to } \ell(\Lambda, \Psi) \leq \ell^* \quad (3)$$

ここで、 ℓ^* は定数である。 ℓ^* が大きいほどスパースな解が得られる。

ここで、定数 ℓ^* は、データへの当てはまりの良さとスパース性の強さのバランスを制御するパラメータである。 $\ell^* = \hat{\ell}$ の場合、(3) 式の解は最尤推定値と一致し、 ℓ^* が大きくなるにつれ、 Λ の推定

値はスパースになる．そして，(3) 式の問題は，次の関数 $\ell_\rho(\Lambda, \Psi)$ を最小化することで得られる [1]．

$$\ell_\rho(\Lambda, \Psi) = \ell(\Lambda, \Psi) + \rho Q(\Lambda), \quad (4)$$

これはまさに， $\rho > 0$ を正則化パラメータとした正則化法である．

4 二つの $Q(\Lambda)$

ここで， $Q(\Lambda)$ は，因子を解釈しやすくするような正則化項である．以下，2 つの正則化項を用いる．

4.1 MCP (Minimax Concave Penalty)

MCP[2] は，Lasso よりもスパースな解が得られ，以下で与えられる．

$$\rho P(|\theta|; \rho; \gamma) = \rho \left(|\theta| - \frac{\theta^2}{2\rho\gamma} \right) I(|\theta| < \rho\gamma) + \frac{\rho^2\gamma}{2} I(|\theta| \geq \rho\gamma) \quad (5)$$

ここで， $\gamma > 1$ はチューニングパラメータである．一般のスパース推定を行うときに MCP は便利である．

4.2 Prenet (Product-based elastic net) ペナルティ

Prenet ペナルティ [3] は以下で与えられる．

$$P(\Lambda) = \sum_{i=1}^p \sum_{j=1}^{m-1} \sum_{k>j} \left\{ \gamma |\lambda_{ij} \lambda_{ik}| + \frac{1}{2} (1 - \gamma) (\lambda_{ij} \lambda_{ik})^2 \right\} \quad (6)$$

$\gamma \in (0, 1]$ はチューニングパラメータである．Prenet ペナルティは quartimin 基準の一般化であり， ρ が十分大きいときに完全単純構造を推定できる．そのため，単純構造推定の場合には，MCP よりも精度が良い場合がある．

これらの提案手法は高次元データに適用でき，これまで因子分析があまり適用されなかったデータ，たとえば fMRI データに応用することができる．

参考文献

- [1] Hirose, K. & Yamamoto, M.: Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing* **25**(5), 863–875. (2015)
- [2] Zhang, C. H.: Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38**(2), 894–942. (2010)
- [3] Hirose, K. & Terada, Y. (2023). Sparse and simple structure estimation via prenet penalization. *Psychometrika*, **88**(4), 1381–1406.