# 深層状態空間モデルによるfMRI画像を用いた精神疾患診断

Mental Disorder Diagnosis using fMRI Images by Deep State-Space Model

| 草野航希        | 田代哲生           | 松原崇               |
|-------------|----------------|-------------------|
| Koki Kusano | Tetsuo Tashiro | Takashi Matsubara |

神戸大学 大学院システム情報学研究科

Graduate School of System Informatics, Kobe university

Machine learning-based accurate diagnosis of mental disorders is expected to support finding their biomarkers and understanding their underlying mechanism. Recent studies employed dynamic and generative models due to the time-varying nature of the brain activities. Though it is difficult to extract complex features due to the simpleness of the model. In this paper, we model fMRI data using dynamic and deep generative model. The proposed deep state-space model is flexible, dynamic and generative(interpretable). Hence it can extract complex feature, capture time-varying nature of the brain activities and identifies brain regions related to the disorders.

# 1. はじめに

早期の正確な神経疾患の診断は,適切な治療や新たな治療法 の評価に役立つ.そのためには,なるべく診断手順において, 疾患のバイオマーカーを特定し,内部メカニズムを解明するべ きで,言い換えれば説明可能であるべきである(表1).また, 疾患は,脳活動や機能的接続(functional connectivities; FC) と呼ばれる脳活動の相互作用を変化させるが,FC はそもそも 経時的に変化するものである[Liu 13,Li 15].それゆえ,時間 変化する信号を捉えるためには,診断手順は動的であるべきで ある(表1).

そういった理由で, fMRI (functional magnetic resonance imaging) は多くの脳画像技術において注目を集めてい る [Atkinson 01]. fMRI は脳領域ごとの活動を視覚化し, 統合 失調症 [Castro 16, Plis 14] などの精神疾患の分析に使用され, また,多くの機械学習手法により分析されてきた(表1).脳領 域ごとの活動間のピアソン相関係数 (PCC) に基づいた特徴抽 出は,疾患によって変化した静的な FCs を検出している [Shen 10, Yahata 16]. 隠れマルコフモデル (hidden Markov model; HMM) は確率的グラフィカルモデルで、脳活動の根底にある ダイナミクスを推測することができる [Suk 16]. 機械学習の中 でも、有力な手法として深層ニューラルネットワーク (DNN) が注目を集めている [Bengio 12, Schmidhuber 15]. DNN は, 高度な特徴を自動で学習することで様々なタスクで成果を挙げ てきた. DNN を再帰的に接続した, LSTM と呼ばれるモデル が, fMRI データの分析に有効であった [Dvornek 17]. LSTM は DNN の柔軟性を享受し (表 1), fMRI データを効果的に扱 う事ができる.だがしかし、複雑な識別機であるため、PCC や HMM とは異なり、疾患の内部メカニズムを抽出するのは 難しい. つまりは, 説明可能ではない. DNN の柔軟性と確率 的グラフィカルモデルの分析的性質を活用するために, deep generative model (DGM) が提案され [Kingma 14], 静的な FC に限定した DGM を fMRI データに適用された [Tashiro 17].

ここで fMRI データ分析に提案する深層状態空間モデル (deep state-space model; DSSM) は、比較手法らの全ての利点を備 えている (表 1). DSSM は、DNN を使った連続的な fMRI デー

表 1: Conceptual Comparison Between Our Proposed Model and Comparative Models

上原邦昭

Kuniaki Uehara

| Model           | 説明可能性        | 動的           | 柔軟性          |
|-----------------|--------------|--------------|--------------|
| PCC             | $\checkmark$ |              |              |
| HMM             | $\checkmark$ | $\checkmark$ |              |
| LSTM            |              | $\checkmark$ | $\checkmark$ |
| DGM             | $\checkmark$ |              | $\checkmark$ |
| DSSM (proposed) | $\checkmark$ | $\checkmark$ | $\checkmark$ |

タの状態空間モデルで, deep Kalman filter [Krishnan 17] の ように DGM を動的に拡張したものとみなせる. deep Kalman filter と異なり, DSSM は患者ごとに診断結果を持っている. DSSM は患者の状態を仮定し, fMRI 画像に尤度を割り当てる ことで, ベイズ則を用いることで診断する.

本研究では,統合失調症と双極性障害の fMRI データを用 いて実験を行い,DSSM を評価した.この実験では,提案し たDSSM が,従来手法 [Shen 10, Yahata 16, Suk 16, Dvornek 17, Tashiro 17] よりも良い精度の診断を達成することを示す.

## 2. Deep State-Space Model

いくつかの先行研究では, 確率的グラフィカルモデルに基 づく診断手順を共通している [Suk 16, Tashiro 17, Matsubara 17]. 確率的グラフィカルモデルは, 可視変数 x, y, と潜在変 数 z の同時確率  $p_{\theta}(x, y, z)$  で定義される. ここでは, 可視変 数  $x = (x_1, x_2, x_3 \dots, x_T)$ は, 連続的な fMRI 画像を,  $x_t$ は, 時刻 t でのスキャンをそれぞれ表す. y は 2 値データで, 患 者の状態 (y = 1 は陽性, y = 0 は陰性)を表す. 潜在変数  $z = (z_1, z_2, z_3, \dots, z_T)$ は, 状態 y が与えられたときに fMRI 画像 x を生成する内部メカニズムとなるような内部状態を表 す. 内部状態 z は, 感情や思考などの, 患者の時間的状態を 表しているとみなされる. 内部状態 z は, 不可観測なので以 下のように周辺化して扱う.

$$p_{\theta}(\boldsymbol{x}, y) = \int_{\boldsymbol{z}} p_{\theta}(\boldsymbol{x}, y, \boldsymbol{z})$$

確率的グラフィカルモデルは、サンプル (x,y) が与えられ

連絡先: 神戸大学 大学院システム情報学研究科 計算科学専攻 E-mail: kusano@ai.cs.kobe-u.ac.jp



図 1: 提案した DSSM  $p_{\theta}$  のグラフィカルモデル.  $\boldsymbol{x} = (x_1, x_2, x_3, ...)$  は連続した fMRI 画像を, y は状態を,  $\boldsymbol{z} = (z_1, z_2, z_3, ...)$  は内部状態を, それぞれ表す.

たときにモデル  $p_{\theta}$  から生成される対数尤度  $\log p_{\theta}(\boldsymbol{x}, y)$  を最 大化するように訓練される.まだ y を与えられてない患者  $\boldsymbol{x}$ について考えてみると、ベイズ則を使用して、患者の状態 y の 事後確率を取得できる.

$$p_{\theta}(y|\boldsymbol{x}) = \frac{p_{\theta}(\boldsymbol{x}, y)}{\sum_{y' \in \{0,1\}} p_{\theta}(\boldsymbol{x}, y')} \propto p_{\theta}(\boldsymbol{x}, y)$$

患者の fMRI 画像 x について,状態を仮定して (y = 0, 1) 尤 度を比較することで診断する.患者の fMRI 画像 x について,  $p_{\theta}(x, y = 1) > p_{\theta}(x, y = 0)$  なら y = 1,  $p_{\theta}(x, y = 0) > p_{\theta}(x, y = 1)$  なら y = 0 と診断する.

DSSM は連続した fMRI 画像のあらゆるフォーマットに対し て適用可能である.そのため,各画像  $x_t$  は、3D、2D, k-space 画像、ボクセルのベクトル、関心領域 (region of interest; ROI) の特徴ベクトルや他の抽出した特徴でもよい.

計算リソースの制限や、過学習の回避のため、多くの先行研究が ROI の特徴ベクトルを扱ってきた [Shen 10, Suk 16, Plis 14, Yahata 16, Castro 16, Dvornek 17, Tashiro 17, Matsubara 17]. 疾患に関係する ROI を特定することは、fMRI データに関する研究の主目的の一つである。そのため、これ以降は fMRI 画像 x として、連続した ROI の特徴ベクトルを使用する.

図 1,2において,黒丸は可視変数,白丸が潜在変数を表す.時 刻 t における内部状態を  $z_t$  と表す.状態 y が与えられたときに, 内部状態  $z_{t-1}$  が次の内部状態  $z_t$  に推移する (transition model  $p_{\theta}(z_t|z_{t-1}, y)$ ).内部状態  $z_t$  が与えられたとき, fMRI 画像  $x_t$ を生成する (emission model  $p_{\theta}(x_t|z_t)$ ). これは AutoEncoder における decoder に一致する.同時確率  $p_{\theta}(x, y, z)$  は以下の ように表される.

$$p_{\theta}(\boldsymbol{x}, y, \boldsymbol{z}) = p_{\theta}(y) p_{\theta}(z_1|y) \prod_{t=2}^{T} p_{\theta}(z_t|z_{t-1}, y) p_{\theta}(x_t|z_t) \quad (1)$$

transition model  $p_{\theta}(z_t|z_{t-1}, y)$ と emission model  $p_{\theta}(x_t|z_t)$ は DNN を使って実装することで、高い柔軟性をもたせた. 変分推論 [Kingma 14, Tashiro 17, Matsubara 17, Krishnan 17] を用いるため、 [Krishnan 17] のような推論モデル  $q_{\phi}$ を設計し、図 2 に示すように以下の平均場近似を採用した.



図 2: 推論  $q_{\phi}$  のグラフィカルモデル.  $\boldsymbol{x} = (x_1, x_2, x_3, \dots)$ は 連続した fMRI 画像を, y は状態を,  $\boldsymbol{z} = (z_1, z_2, z_3, \dots)$  は内 部状態を,それぞれ表す.



図 3: DSSM の実装イメージ図.

$$q_{\phi}(\boldsymbol{z}|\boldsymbol{x}, y) = \prod_{t=1}^{T} q_{\phi}(z_t|\boldsymbol{x}, y)$$
(2)

これらのモデルを使うことで、モデルエビデンスと呼ばれ る可視変数 x, yの対数尤度  $\log p_{\theta}(x, y)$  は以下のように表さ れる.

$$\log p_{\theta}(\boldsymbol{x}, y) \geq \mathbb{E}_{q_{\phi}(\boldsymbol{z}|\boldsymbol{x}, y)} \left[ \log \frac{p_{\theta}(\boldsymbol{x}, y, \boldsymbol{z})}{q_{\phi}(\boldsymbol{z}|\boldsymbol{x}, y)} \right]$$
$$= \sum_{t=1}^{T} \mathbb{E}_{q_{\phi}(z_{t}|\boldsymbol{x}, y)} \left[ \log p_{\theta}(x_{t}|z_{t}) \right]$$
$$- \sum_{t=2}^{T} \mathbb{E}_{q_{\phi}(z_{t-1}|\boldsymbol{x}, y)} \left[ D_{KL}(q_{\phi}(z_{t}|\boldsymbol{x}, y) || p_{\theta}(z_{t}|z_{t-1}, y)) \right]$$
$$- D_{KL}(q_{\phi}(z_{1}|\boldsymbol{x}, y) || p_{\theta}(z_{1}|y)) + \log p_{\theta}(y)$$
$$=: \mathcal{L}_{g}(\boldsymbol{x}, y).$$
(3)

モデルエビデンスの下界  $\mathcal{L}_g(\boldsymbol{x}, y)$  を,目的関数とし,最大化する.  $\mathcal{L}_g(\boldsymbol{x}, y)$ は,  $\log p_{\theta}(\boldsymbol{x}, y)$ に収束すると考えられている [Blei

| 表 2: Diagnosis Accuracies         |               |               |         |               |  |  |  |
|-----------------------------------|---------------|---------------|---------|---------------|--|--|--|
|                                   | Schizophrenia |               | Bipolar |               |  |  |  |
| Model                             | BACC          | $\mathbf{F1}$ | BACC    | $\mathbf{F1}$ |  |  |  |
| PCC+SCCA+SLR [Yahata 16]          | 0.646         | 0.462         | 0.519   | 0.378         |  |  |  |
| PCC+Kendall+LLE+c-means [Shen 10] | 0.670         | 0.544         | 0.626   | 0.480         |  |  |  |
| LSTM [Dvornek 17]                 | 0.673         | 0.522         | 0.577   | 0.345         |  |  |  |
| AE+HMM [Suk 16]                   | 0.579         | 0.429         | 0.589   | 0.417         |  |  |  |
| DGM [Tashiro 17]                  | 0.707         | 0.571         | 0.633   | 0.445         |  |  |  |
| DSSM (proposed)                   | 0.728         | 0.601         | 0.633   | 0.481         |  |  |  |

 ${\bf \bar{z}}$  3: Regions with Top 3 and Bottom 3 Contribution Weights for Diagnosis.

| Schizophrenia       |                           | Bipol               | Bipolar                   |  |  |
|---------------------|---------------------------|---------------------|---------------------------|--|--|
| ROI                 | weight $(\times 10^{-4})$ | ROI                 | weight $(\times 10^{-4})$ |  |  |
| Postcentral_R       | 3.08                      | Temporal_Inf_R      | 1.49                      |  |  |
| Precentral_R        | 3.00                      | $Cingulum\_Post\_R$ | 1.14                      |  |  |
| Postcentral_L       | 2.81                      | Postcentral_R       | 1.01                      |  |  |
| :                   | ÷                         | ÷                   | :                         |  |  |
| Putamen_L           | 0.04                      | $Cerebelum_7b_L$    | -0.26                     |  |  |
| Temporal_Pole_Sup_R | 0.02                      | $Cerebelum_Crus1_L$ | -0.28                     |  |  |
| Vermis_1_2          | -0.02                     | $Cingulum\_Ant\_R$  | -0.32                     |  |  |

17, Bishop 06] ので、 $\log p_{\theta}(x, y)$ の代わりに使用される. 患者の状態 yの事後分布  $p_{\theta}(y|x)$  は以下のように近似される.

$$p_{\theta}(y|\boldsymbol{x}) = \frac{p_{\theta}(\boldsymbol{x}, y)}{\sum_{y' \in \{0,1\}} p_{\theta}(\boldsymbol{x}, y')} \\ \approx \frac{\exp \mathcal{L}_g(\boldsymbol{x}, y)}{\sum_{y' \in \{0,1\}} \exp \mathcal{L}_g(\boldsymbol{x}, y')}$$
(4)
$$= \exp \mathcal{L}_d(\boldsymbol{x}, y)$$

加えて,  $\mathcal{L}_d(\boldsymbol{x}, \boldsymbol{y})$ は, 状態  $\boldsymbol{y}$  の対数尤度  $\log p_{\theta}(\boldsymbol{y}|\boldsymbol{x})$  の近 似であり, クラス分類を進めるような代わりの目的関数にな る [Lasserre 06]. 以下のように定数  $\omega \in [0, 1]$  を使って,目的 関数のバランスをとる.

$$\mathcal{L}(\boldsymbol{x}, y) = (1 - \omega)\mathcal{L}_g(\boldsymbol{x}, y) + \omega\mathcal{L}_d(\boldsymbol{x}, y)$$
(5)

実装イメージは、図3に示す.

[Tashiro 17] のように, emission model  $p_{\theta}(x_t|z_t)$  は, 隠れ 層が 2 層のニューラルネットワークを採用した. 基本的に各 隠れ層には, layer normalization を行い [Ba 16], 活性化関数 には ReLU を用いた [Nair 10]. emission model  $p_{\theta}(x_t|z_t)$  は, reparameterization trick [Kingma 14] を使って,  $x_t$ の事後分 布  $\mathcal{N}(\mu_{z_t}, \operatorname{diag}(\sigma_{z_t}^2))$  を出力する.

transition model  $p_{\theta}(z_t|z_{t-1}, y)$ には, [Krishnan 17]の提 案した DNN ベースの gated transition を用いた. これによ り, gated recurrent unit(GRU) と同様に, 各次元とステッ プに対して線形, 非線形な遷移を調整できる柔軟性をもたせて いる. ここでは [Krishnan 17] とは異なり, 状態 y と前内部状 態  $z_{t-1}$ の組を扱う実装にし, 事後分布  $p_{\theta}(z_t|z_{t-1}, y)$ のパラ メータを出力した. 推論モデル  $q_{\phi}(\boldsymbol{z}|\boldsymbol{x}, y)$  は, bidirectional LSTM [Hochreiter 97, Schuster 97] で実装した. bidirectional LSTM は, 順方 向と逆方向の出力をもち,出力された順方向と逆方向の分布  $q_{\phi}^{(f)}(z_t|\boldsymbol{x}, y), q_{\phi}^{(b)}(z_t|\boldsymbol{x}, y)$ から計算し,事後分布 $q_{\phi}(z_t|\boldsymbol{x}, y) \propto$  $q_{\phi}^{(f)}(z_t|\boldsymbol{x}, y) \times q_{\phi}^{(b)}(z_t|\boldsymbol{x}, y)$ を得る.これにより, Eq. (3) の Kullback-Leibler 距離  $D_{KL}(q_{\phi}(z_t|\boldsymbol{x}, y)||p_{\theta}(z_t|z_{t-1}, y))$ が計 算できる,

## 3. 実験

本実験で用いたデータセットは、OpenfMRI データベース の ds000030 (https://openfmri.org/dataset/ds000030/) から得た,前処理済み [Esteban 18](https://github. com/poldracklab/fmriprep) データである.このデータに scrubbing を行った後に,automated anatomical labeling (AAL) [Tzourio-Mazoyer 02] を使用し、各 ROI 領域におけ るボクセル強度を平均化することで,fMRI 画像を 116 の ROI に分割し,ROI の特徴ベクトルを得た.最後に各時間の ROI 強度を.0.01Hz から 0.1Hz の周波数帯になるようにバンドパ スフィルタを通して,それを平均0,分散1 になるように正規 化を施した.それにより,最終的に健常者 117 人,統合失調症 患者 48 人,双極性障害患者 46 人を使用した.

DSSM の各隠れ層は  $n_h$  個の隠れユニットを持ち,内部状態  $z_t$  は,  $n_z$  次元である.各重みは  $\mathcal{N}(0, 0.02^2)$  からのサンプル, 各バイアスは 0 で初期化した.入力層には,dropout 率 p で dropout を行った.3つのモデル全ての学習には,学習率は  $10^{-4}$ その他は初期パラメータの Adam アルゴリズム [Kingma 15] を使用した. $n_h \in \{100, 200, 400\}, n_z \in \{5, 10, 20, 50, 100\},$   $p \in \{0.0, 0.5\}, \omega \in \{0.0, 0.9, 0.95, 0.99\}$ の範囲でハイパーパ ラメータを探索した.また,データのクラスの偏りを調整する ために oversampling を行って学習した.比較手法と,その探 索パラメータは [Tashiro 17] と同様である.

## 4. 結果

実験結果を,表2に示す.なお,10-fold cross-validation を5 回行った平均値を用いて精度を評価している.また,データセッ トに偏りがあることを考慮して,accuracy ではなく balanced accuracy(BACC)で評価する.この結果は,表1に示した3 つの特性によるものだと考えられる.更に詳しい解析は今後明 らかにしていきたい.

クラスラベルを変えたときの再構成誤差の変化の大きさを 計算することで、診断に特に寄与した ROI を調べることが できる. 具体的な計算方法は [Matsubara 17] と同様である. 各 ROI の寄与度を求め、上位 3 位と下位 3 位の ROI を表 3 に示す. 統合失調症の場合, 上位3位の ROI の寄与度は大 きな値で関係があることを,下位3位においてはほぼ0で関 係がないことをそれぞれ示唆している. 統合失調症に対する post-central と pre-central の関係は、いくつかの研究で言及 されてきた [Tanskanen 10] が, 統合失調症の研究の中では それほど一般的ではない [Glahn 08]. 双極性障害の場合,上 位3位のROIの寄与度はそれほど大きくなく,偶然程度のレ ベルとはいえ下位3位は0もしくはわずかに負となった.こ の結果は、識別がまだ信頼に足りない可能性を示唆している. 生成モデルを使って ROI の特徴の再構成誤差 (負の対数尤度  $-\mathbb{E}_{q_{\phi}(z_t|\boldsymbol{x},y)}[\log p_{\theta}(x_t|z_t)])$ に基づいて診断したが,これは, 推論された内部状態と生成された内部状態 z の間の誤差であ る Kullback-Leibler 距離  $D_{KL}(q_{\phi}(z_t|\boldsymbol{x}, y)||p_{\theta}(z_t|z_{t-1}, y))$  に も依存する. DSSM の transition や推論モデルが豊かなため に、内部状態が ROI 方向の特徴に対して優位をもつことが考 えられる.内部状態 z に関する詳しい分析は今後明らかにし ていきたい.

## 5. 結論

本研究では、精神疾患の診断のために構成した fMRI 画像 の深層状態空間モデル (deep state-space model; DSSM) を提 案した.DSSM は、動的な機能的接続を考慮して、fMRI 画像 と疾患と時間変化する内部状態の同時確率をモデル化したもの である.内部状態の動的モデルにより、DSSM は従来手法及 び、比較手法よりも高い診断精度を示した.DSSM は疾患に 関係のある脳状態を示唆し、潜在的な治療法の開発に役立つ可 能性を示している.

本研究は総務省 SCOPE(受付番号 172107101) の委託を受けて行われた.

# 参考文献

- [Atkinson 01] Atkinson, A. J., et al.: Clinical Pharmacology and Therapeutics, vol. 69, no. 3, pp. 89–95, (2001)
- [Ba 16] Ba, J. L. et al.: Layer Normalization (2016)
- [Bengio 12] Bengio, Y. et al.: IEEE Transactions on Pattern Analysis and Machine Intelligence, No. 1993, pp. 1–30 (2012)
- [Bishop 06] Bishop, C. C. M. C.: Pattern recognition and machine learning, in *Springer*, Vol. 4, p. 738 (2006)

- [Blei 17] Blei, D. M. et al.: Journal of the American Statistical Association, Vol. 112, No. 518, pp. 859–877 (2017)
- [Castro 16] Castro, E. et al.: IEEE Transaction on Medical Imaging, Vol. 35, No. 7, pp. 1729–1740 (2016)
- [Dvornek 17] Dvornek, N. C. et al.: International Workshop on Machine Learning in Medical Imaging, Vol. 10541, pp. 362–370 (2017)
- [Esteban 18] Esteban, O. et al.: bioRxiv, pp. 1–20 (2018)
- [Glahn 08] Glahn, D. C. et al.: Biological Psychiatry, Vol. 64, No. 9, pp. 774–781 (2008)
- [Hochreiter 97] Hochreiter, S. et al.:Neural Computation, Vol. 9, No. 8, pp. 1735–1780 (1997)
- [Kingma 14] Kingma, D. P. et al.: International Conference on Learning Representations (2014)
- [Kingma 15] Kingma, D. P. et al.: International Conference on Learning Representations, pp. 1–15 (2015)
- [Krishnan 17] Krishnan, R. G. et al.: in AAAI Conference on Artificial Intelligence, pp. 1–13 (2017)
- [Lasserre 06] Lasserre, J. A. et al.: in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 87–94, IEEE (2006)
- [Li 15] Li, X. et al.: BioMed Research International, Vol. 2015, (2015)
- [Liu 13] Liu, X. et al.: Proceedings of the National Academy of Sciences, Vol. 110, No. 11, pp. 4392–4397 (2013)
- [Matsubara 17] Matsubara, T. et al.: arXiv (2017)
- [Nair 10] Nair, V. et al.: in International Conference on Machine Learning, pp. 807–814 (2010)
- [Plis 14] Plis, S. M. et al.: Frontiers in Neuroscience, Vol. 8, No. Aug., pp. 1–11 (2014)
- [Schmidhuber 15] Schmidhuber, J.: Neural Networks, Vol. 61, pp. 85–117 (2015)
- [Schuster 97] Schuster, M. et al.: IEEE Transactions on Signal Processing, Vol. 45, No. 11, pp. 2673–2681 (1997)
- [Shen 10] Shen, H. et al.: NeuroImage, Vol. 49, No. 4, pp. 3110–3121 (2010)
- [Suk 16] Suk, H. I. et al.: NeuroImage, Vol. 129, pp. 292– 307 (2016)
- [Tanskanen 10] Tanskanen, P. et al.: Schizophrenia Bulletin, Vol. 36, No. 4, pp. 766–777 (2010)
- [Tashiro 17] Tashiro, T. et al.: NOLTA, pp. 700-703 (2017)
- [Tzourio-Mazoyer 02] Tzourio-Mazoyer, N. et al.: NeuroImage, Vol. 15, No. 1, pp. 273–289 (2002)
- [Yahata 16] Yahata, N. et al.: Nature Communications, Vol. 7, No. 7, p. 11254 (2016)