

神経活動と行動指標の不確実性を統一して扱う結合潜在過程モデル A Joint Latent Process Model of Temporal Uncertainty in Brain Activity and Behavior

浅川 伸一[†], 竹市 博臣[‡], 鈴木 航[‡], 山下 和香代^{*}

Shinichi Asakawa, Hiroshige Takeichi, Wataru Suzuki, Wakayo Yamashita

[†] 東京女子大学情報処理センター, [‡] 理化学研究所生命医科学研究センター予測医学特別プロジェクト, ^{*} 鹿児島大学工学部

[†]Tokyo Woman's Christian University, [‡]PMSP, CIMS, RIKEN, ^{*}Faculty of Engineering, Kagoshima University
asakawa@ieee.org

概要

神経活動と行動とを統一的に説明する計算モデルの構築は認知科学・神経科学における中心課題であるが、時間的不確実性はなお表現学習、神経モデリング、行動読み出しの各段階で断片的に扱われることが多い。本研究では、人間の場面分類課題における EEG データを用い、この断片が解釈上の失敗を引き起こすことを論じる。すなわち、既存モデルは時間的パターンの再現や行動予測は可能である一方、神経データ変動と行動データ変動とを首尾一貫して関連付けることができない。我々は、この問題がアルゴリズムではなくモデル構造に起因すると捉え、不確実性を単一の潜在過程に集約する確率的連結フレームワークを提案する。潜在過程にはガウス過程事前分布を課し、刺激類似性はノンパラメトリックな連結関数を通じて潜在状態の初期化に反映される。さらに、神経活動データと行動データとに個別の変分因子を導入せず、結合潜在過程に対する統一事後表現を維持することで、不確実性の断片を推論段階でも回避する。この構造の下では、不確実性の統一により EEG 時系列変動と行動データとの対応が回復する一方、時間構造を持つが断片されたベースラインではこの対応は得られないと予測される。以上より、神経活動データと行動データにおける時間的不確実性を統一的に扱う試みはモデリングにおける構造的条件であると論じる。

キーワード: Brain-behavior linking, Gaussian process latent models, Neural time series modeling, Interpretability in cognitive models

1. はじめに

神経活動データと行動データとを統一的に説明する計算モデルの構築は、認知科学・神経科学における中心課題である。とくに近年では、神経活動時系列データに対して時間構造を明示的に導入する試みが広く行われている。しかし、時間的不確実性はなお表現学

習、神経モデリング、行動読み出しの各段階で断片的に扱われることが多く、その結果として神経活動データ変動と行動データ変動との関係が一貫して説明されないという問題が残されている。

図 1 は、場面分類課題における EEG 波形の典型例を示す (Karapetian et al., 2023)。後頭・頭頂部電極において観測される応答は、試行間で高い時間的一貫性を示す一方、カテゴリ差は単一のピークとしてではなく、時間全体にわたる分散的な変調として現れる。このような構造は、時間情報を潰した静的表現では捉えられず、また時間構造を導入したモデルであっても、時間的不確実性の扱いが断片されている限り、神経活動と反応時間との対応を適切に説明することができない。

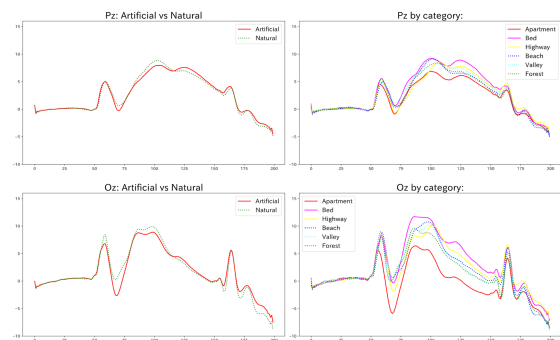


図 1 場面分類中の EEG 反応における時間的構造と分散型カテゴリ効果。

本研究では、この問題をアルゴリズムの選択ではなく、モデル構造の問題として捉える。すなわち、神経活動と行動を結びつけるためには、時間的不確実性を単一の潜在過程に集約する必要があると考える。この立場から、我々はガウス過程を用いた潜在時間モデルと、刺激特徴から潜在状態への連結関数とを組み合わせた確率的フレームワークを提案する。本モデルでは、EEG と行動は共通の潜在過程から生成されるだけでなく、推論の段階においても結合事後分布を通じて

結びつけられるため、時間的不確実性が両者の間で一貫して伝播する。

実証的には、時間情報を潰したモデルおよび時間構造を導入したが不確実性が分断されたモデルと比較することで、提案手法の役割を明確化する。その結果、時間的不確実性を統一的に扱うことで初めて、EEG 時系列変動と行動データ変動の対応が回復することを示す。以上より、本研究は時間的不確実性の統一が神経活動データと行動データとを統一的に扱いうる構造的条件であることを示す。

2. 関連研究

神経活動データと行動データとの関係を記述する連結モデルは、潜在的な認知状態を介して両者を結びつける枠組みとして発展してきた。従来研究では、神経活動指標と行動指標との間に比較的単純な対応関係を仮定することが多かったが、近年では非線形かつ非パラメトリックな連結関数を用いて、刺激・神経活動・行動の対応を柔軟に表現する試みが進んでいる。しかし、その多くは試行単位の集約特徴量に依存しており、神経時系列に内在する時間的不確実性そのものを連結の中核に据えてはいない。本研究は、連結モデルの立場を踏襲しつつ、神経活動データと行動データの関係成り立たせるためには、時間的不確実性を結合潜在過程として扱う必要があるという点を明示化する。

一方、神経時系列の時間構造を扱うモデルとしては、状態空間モデル、再帰型ニューラルネットワーク、潜在力学系モデル、および潜在変数にガウス過程事前分布を課すモデル (Lawrence, 2005) などが提案されてきた。これらのモデルは、EEG や神経活動の滑らかな変化を表現し、再構成や予測の性能向上に寄与している。しかし、多くの場合、時間構造は神経信号の表現や復元のために導入されており、行動指標とは別の段階で扱われる。そのため、神経側で捉えられた時間変動が、反応時間のような行動変動とどのように対応するのかは、依然として明確ではない。本研究は、時間構造を神経信号の記述手段としてではなく、脳-行動関係を記述するための結合制約として位置づける点でこれらと異なる。

不確実性を明示的に扱う確率的モデルは、神経ノイズ、試行間変動、表象の曖昧さを定量化するための有力な手段である。しかし、既存研究では、不確実性は神経観測、潜在表現、行動読み出しの各段階に個別に導入されることが多く、結果として「どの不確実性がどの変動を生んでいるのか」が曖昧になりやすい。とくに脳活動と行動を同時に扱う場合、不確実性が分断

されていると、神経変動と行動変動の対応は適合されるだけで説明されない。本研究は、この点を構造的問題として捉え、時間的不確実性を単一の潜在過程に集約することで、神経活動と行動との対応を解釈可能にすることを目指す。

3. 方法

前節で述べたように、本研究では時間的不確実性の分断が脳-行動関係の解釈を妨げる要因であると考えられる。そこで本研究では、時間的不確実性を単一の潜在過程に集約する確率的フレームワークを導入する。本節では、この枠組みを構成する潜在時間過程、連結関数、神経データモデル、行動データモデル、および推論と最適化を順に定式化する。

3.1 潜在時間過程

各試行に対して、時刻 $t = 1, \dots, T$ における潜在状態 $\mathbf{z}_t \in \mathbb{R}^D$ を導入する。この潜在過程は、図 1 に示されるような滑らかで一貫した EEG 波形構造を反映するものとし、その時間構造はガウス過程事前分布によって与える：

$$\mathbf{z}_{1:T}^{(d)} \sim \mathcal{N}(\mathbf{0}, K_{\text{time}}) \quad (1)$$

ここで K_{time} は時間カーネルにより定まる共分散行列であり、時間相関と時間的不確実性はすべてこの潜在過程に集約される。

3.2 連結関数

刺激特徴 $\mathbf{s}_n$ は潜在過程の初期状態に作用すると仮定し、

$$\mathbf{z}_1 = g(\mathbf{s}_n), \quad g \sim \mathcal{GP}(\mathbf{0}, k_{\text{stim}}) \quad (2)$$

とする。ここで連結関数 g は刺激空間から潜在状態への対応を与え、表象類似性に関する不確実性を潜在過程に導入する。

3.3 神経データモデル

EEG 観測 $\mathbf{e}_t \in \mathbb{R}^C$ は潜在状態の線形射影として次式とする：

$$\mathbf{e}_t = W\mathbf{z}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_e) \quad (3)$$

時間相関は観測ノイズではなく潜在過程に集約される。

3.4 行動データモデル

行動データは、各試行におけるカテゴリ判断 y_n と反応時間 r_n から構成される。潜在過程の要約のため次式を用いる：

$$\bar{\mathbf{z}}_n = \sum_{t=1}^T \alpha_t \mathbf{z}_t \quad (4)$$

カテゴリ判断については、次式とする：

$$p(y_n | \bar{\mathbf{z}}_n) = \text{softmax}(U\bar{\mathbf{z}}_n) \quad (5)$$

反応時間については、次式の如くモデル化する：

$$\log r_n \sim \mathcal{N}(\mathbf{a}^\top \bar{\mathbf{z}}_n + b, \sigma^2) \quad (6)$$

これにより、カテゴリ判断と反応時間は別個の過程ではなく、共通の潜在過程に基づく行動指標として統一的に記述される。

3.5 推論と最適化

潜在過程 $\mathbf{z}$ および連結関数 g の厳密な事後分布は解析的に求めることができないため、本研究では変分近似を用いる。各試行の潜在過程に対して変分分布 $q(\mathbf{z})$ を導入し、神経データと行動データを同時に条件づけた結合事後分布を近似する。ここで重要なのは、神経成分と行動成分に対して個別の変分因子を導入しないことであり、すべての観測値が同一の潜在過程を条件とすることで、時間的不確実性の統一的な事後表現を強制する点にある。

また、刺激と潜在状態の対応を与える連結関数 g に対しても変分近似を導入する。ただし、関数空間上で個別の変分分布を直接最適化するのではなく、刺激特徴空間において誘導入力を用いた疎なガウス過程近似を採用する。これにより、連結関係に関する不確実性を保持しつつ、初期潜在状態 $\mathbf{z}_{n,1}$ に関する tractable な変分事後分布を得ることができる。

モデル学習は、次式で与えられる ELBO の最大化として定式化される：

$$\mathcal{L} = \mathbb{E}_q[\log p(\mathbf{e}, y, r | \mathbf{z})] - \text{KL}(q(\mathbf{z}) \| p(\mathbf{z})) \quad (7)$$

ここで $p(\mathbf{z})$ は潜在時間過程の事前分布であり、時間的不確実性はこの項にのみ現れる。したがって、神経データと行動データは、結合潜在過程に基づく単一の不確実性を通じて結合される。

本研究の目的は新しい推論アルゴリズムの提案ではなく、推論の段階においても時間的不確実性の統一を維持することにある。この点を検証するために、本研究では時間カーネルを除去して潜在状態を時刻ごとに独立とするアブレーションを導入する。この操作により、神経データの再構成性能がある程度保たれる場合でも、潜在不確実性と反応時間との対応が失われるかどうかを調べる。すなわち、本研究におけるアブレーションは性能比較ではなく、時間的不確実性の統一が構造的条件であるかを検証するための操作である。

3.5.1 最適化

すべてのモデルパラメータと変分パラメータは、ELBO に対する確率的勾配上昇法を用いて同時最適化される。期待値は再パラメータ化勾配を用いて推定され、ミニバッチは試行ごとに形成される。時間カーネルと刺激カーネルのハイパーパラメータは、同じ目的

関数を最大化することで学習され、データによって適切な時間的平滑性と表現の類似性が決定される。

我々の貢献は、新しい推論技術や最適化戦略にあるのではないことを強調する。むしろ、推論は構造的仮説を検証するためのツールとして用いられている。すなわち、時間的不確実性を単一の潜在過程に集中させることで、神経ダイナミクスと行動の関係を説明するのに十分かどうかという仮説である。結果の節で示すように、この仮説は、推論そのものは変更されなくても、時間的不確実性を細分化させるアブレーションによって反証され得る。

4. 結果（予測される結果パターン）

人間の場面分類課題から得られた EEG および行動データを用いて、提案モデルを評価した。すべての結果は、図 2（A～D）の 4 つのパネルに沿って整理されており、実証的な動機付けから、ベースラインモデルの失敗、そして統一モデルの成功へと順を追って示されている。

Figure 2 (schematic). From distributed neural structure to unified brain-behavior uncertainty

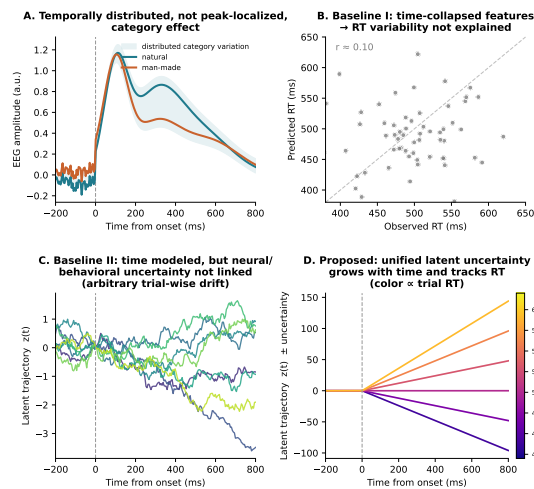


図 2 EEG 波形の時間的一貫性とカテゴリ差の分散的構造を示す概略図（予定）。

本節で示す結果は、提案モデルを実データにフィットして得られた実証結果ではなく、モデルの構造的性質から導かれる予測パターンであり、今後の実データ検証によって確認されるべき仮説である。図 1 は場面分類課題から得られた実際の EEG データを、図 2 はモデル比較によって期待される定性的パターンを模式的に示したものである。

4.1 実データにおける時間的構造と分散したカテゴリ効果

図 1 は、後頭部電極（Pz および Oz）における EEG 応答の全体平均を示している。試行全体を通じて、応

答は初期・中期・後期の成分が滑らかに変化する一貫した時間的プロファイルを示す一方、人工的シーンと自然シーンのカテゴリー関連差は、局所的なピークのシフトではなく時間軸にわたる分散的な変動として現れている。

この観測は2つの制約を導く。第一に、波形に現れる情報を排除することなく、神経活動を静的特徴や時間圧縮特徴だけで適切に要約することはできない。第二に、カテゴリー関連情報が時間的に拡張された過程に埋め込まれているとすれば、行動の変動性を説明する上で時間的不確実性そのものが関与している可能性が示唆される。この観測が、以下で比較する3つのモデル構造（時間圧縮・時間考慮だが断片化・提案モデル）の動機となっている。

4.2 予測される失敗パターンⅠ：時間軸を圧縮したモデル

図2(B)は、EEG反応を時間軸上で圧縮し標準的な回帰・分類モデルで行動結果を予測した場合に予想されるパターンを模式的に示す。このようなモデルはカテゴリー解釈において偶然を上回る精度を達成するが、時間情報を圧縮する時点で図1に見られる分散的な時間構造が失われるため、反応時間の試行間変動を説明する情報も同時に失われると予測される。図2(B)の模式的な散布図が示すように、予測RTと観測RTの対応は弱く、不確実性の推定値も試行ごとの神経応答の違いを反映しないことが見込まれる。

4.3 予測される失敗パターンⅡ：時間を考慮するが不確実性が断片化したモデル

図2(C)は、EEG信号の時間構造を明示的にモデル化するが、神経活動成分と行動成分の間で不確実性を結合しない潜在時系列モデル（例えば個別の変分因子を持つ時間考慮モデル）を想定した場合に予測されるパターンを示す。このようなモデルは滑らかな時間パターンを捉え、時間圧縮モデルより忠実にEEG波形を再構成できると考えられる一方、神経段階と行動段階の間で不確実性が分断されている限り、潜在軌跡は試行間で行動と無関係な恣意的ドリフトを示し、神経表現の不確実性は反応時間の不確実性を反映しないと予測される（図2(C)）。この予測が正しければ、時間構造の導入だけでは脳・行動対応の回復に不十分であることになる。

4.4 提案モデルにより期待される脳・行動対応の回復

図2(D)は、神経学的観測と行動的観測が単一の結合潜在過程を通じて生成される提案モデルにおいて期待されるパターンを示す。式(1)～(7)の構造の下では、潜在過程における不確実性は時間とともに増大

し、EEG予測と行動出力の双方に一貫して伝播するはずである。この構造が妥当であれば、潜在不確実性が大きい試行ほど反応時間がより長く変動しやすくなり、推定された時間的不確実性と行動の遅延との間に単調な関係が生じると予測される。この対応は、4.2・4.3で述べたいずれのベースラインでも得られないと予測される点が、提案モデルの構造的な特徴である。

この予測の必要条件を検証するための操作として、時間的事前分布を時間軸上で独立な事前分布に置き換えるアブレーションを導入する（3.5節）。この操作は再構成の質をある程度維持しつつ、神経活動と行動の変動性との対応関係を破壊すると予測される。もしこの予測が実データで確認されれば、解釈可能性のためには予測性能そのものではなく時間的不確実性の統合が必要であることの証拠となる。

4.5 予測の要約

以上の予測パターンを総合すると、3つの検証可能な仮説が導かれる。第一に、EEGデータにおける時間的構造は分散しており静的要約では捉えきれない（4.1で実データにより確認済み）。第二に、神経的要素と行動的要素にわたる不確実性を統合せずに時間的構造をモデル化すると、脳・行動対応の解釈上の失敗を招くと予測される（4.2・4.3）。第三に、時間的不確実性を共通の潜在過程に集約することで、複雑なダイナミクスや課題固有の機構を追加導入することなく脳・行動間の整合的な関係が回復すると予測される（4.4）。

5. 結論

本研究は、神経活動と行動データとを統一的に説明するためには、時間的不確実性を単一の潜在過程として扱うことが必要であることを示した。時間を潰したモデルおよび時間構造を導入したが不確実性が分断されたモデルでは、神経変動と反応時間変動の対応は回復しなかった。これに対して提案モデルは、結合潜在過程を通じて両者を結びつけ、時間的不確実性の統一が脳・行動モデリングにおける構造的条件であることを明らかにした。

文 献

- Karapetian, A., Boyanova, A., Pandaram, M., Obermayer, K., Kietzmann, T. C., & Cichy, R. M. (2023). Empirically Identifying and Computationally Modeling the Brain-Behavior Relationship for Human Scene Categorization. *Journal of Cognitive Neuroscience*, 35(11), 1879–1897.
- Lawrence, N. D. (2005). Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models. *Journal of Machine Learning Research*, 6, 1783–1816.