

2026年6月12日

報道機関 各位

国立大学法人東北大学

**少ない脳画像データから再現性の高い予測が可能に
- 大規模脳画像データの「主成分」を小規模研究に転用し、
脳行動予測の再現性を向上 -**

【発表のポイント】

- 大脳皮質厚データに主成分分析（PCA）^{（注1）}を行った場合、100人未満のサンプルでは主成分が極めて不安定になることを定量的に証明しました。サンプル間の安定性の94～97%がサンプルサイズで説明されました。
- 大規模データの「主成分ベクトル」^{（注2）}をデータが重複しない小規模サンプルに転用すると、認知能力・性格特性の予測精度が有意に向上しました。
- 採用する主成分数を検討した結果、最適なピーク値があることが示されました。

【概要】

脳MRI研究では、少ない被験者で解析すると結果が再現されないという「再現性危機」が深刻な問題です。特に16万4千次元に及ぶ脳画像データへの主成分分析（PCA）は、サンプルが少ないと主成分ベクトル（固有ベクトル）が安定せず、後続の解析全体の信頼性を損ないます。

東北大学大学院情報科学研究科（兼：加齢医学研究所）の細田千尋准教授らの研究グループは、この問題を解決する実践的な戦略として「大規模データの固有ベクトルを小規模データに転用する」手法を提案・実証しました。

1,113名分の脳画像データ（Human Connectome Project）^{（注3）}を用いた体系的実験の結果、たった100人のデータでも、500人規模データの固有ベクトルを利用するだけで、認知能力・性格特性の脳行動予測精度が大幅に向上することが明らかになりました。

本成果は、限られたリソースしか持たない多くの研究機関に対して、公開大規模データを基盤として活用するという新たな研究設計の指針を提供する可能性を示唆しました。本研究は5月25日に科学誌 Scientific Reports（Nature Portfolio）に掲載されました。

【詳細な説明】

研究の背景

脳神経画像研究における「再現性危機」は世界的な課題です。fMRI 研究では 20~30 人規模の研究が信頼できる結果を出すことはほぼ不可能と言われ、安定した脳行動関連を得るには 1,000 人以上が必要との報告もあります (Marek et al., Nature 2022)。しかし、大規模データ収集はコスト・倫理・設備の面で多くの機関には困難です。

PCA は脳画像の次元削減^(注1)に広く使われますが、サンプルが少ない「高次元低サンプル数 (HDLSS)^(注4)」の場合、主成分分析で得られる結果は、サンプルの違いによってばらつきやすく、不安定になりやすいことが知られています。しかし、その影響の大きさや解決方法は十分に明らかになっていませんでした。

今回の取り組み

本研究では、1,113 人分の脳画像データから、人数の異なる小さなサンプルデータを多数作り、分析結果がどの程度安定するか (同じになるか) を調べました。その結果、少人数サンプルデータでは、脳のどの特徴が重要かという結果が分析ごとに異なりばらつきやすい一方、人数が増えるほど同じ特徴が安定して見つかることが分かりました。実際に、結果のばらつきの 94~97% は、サンプルデータの人数の違いで説明されました。100 人未満では安定した特徴がほとんど得られませんでした。375 人以上では多くの特徴が安定して抽出されました。

次に、大規模データから得られた安定した特徴を、小規模データの解析に活用できるかを検証しました。その結果、100 人のデータだけで分析した場合に比べ、500 人のデータから得た特徴を利用して 100 人のデータを解析した場合には、認知能力や性格特性をより高い精度で予測できました。特に流動性知能や作業記憶では、大規模データで分析した場合に近い精度が得られました。また、使う特徴の数は多ければよいわけではなく、今回のデータでは約 30 個を用いた場合に予測精度が最も高くなりました。(下図参照)。

今後の展開

本研究で提案した手法は、これまで課題とされてきた脳画像研究の「再現性問題」を改善する新たなアプローチです。公開されている大規模データから得られた安定した特徴を活用することで、少人数しか集められない研究でも、より信頼性の高い結果を得られる可能性があります。今後は、患者や子どもなど大規模なデータ収集が難しい研究への応用を進め、脳の個人差に基づく医療や教育の発展につなげることを目指します。

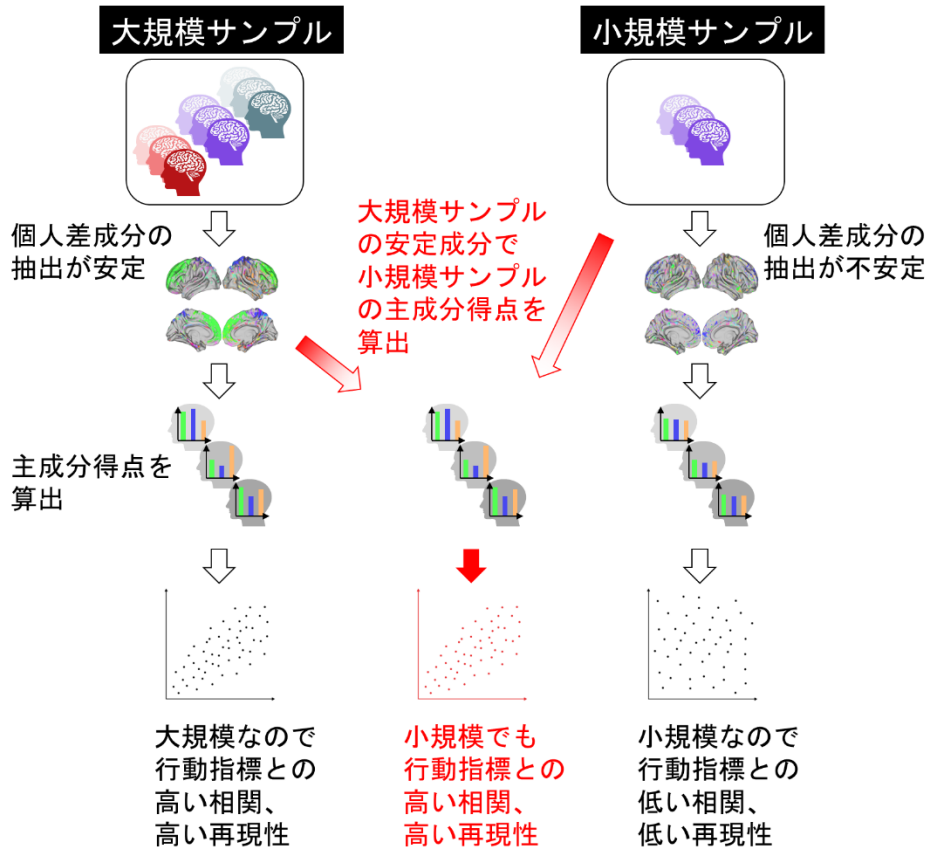


図 1. 新手法の模式図。大規模データから抽出した統計的に安定した成分を、統計的に不安定な小規模データに適用することによって、小規模データの解析の統計的安定性を高めることができる。

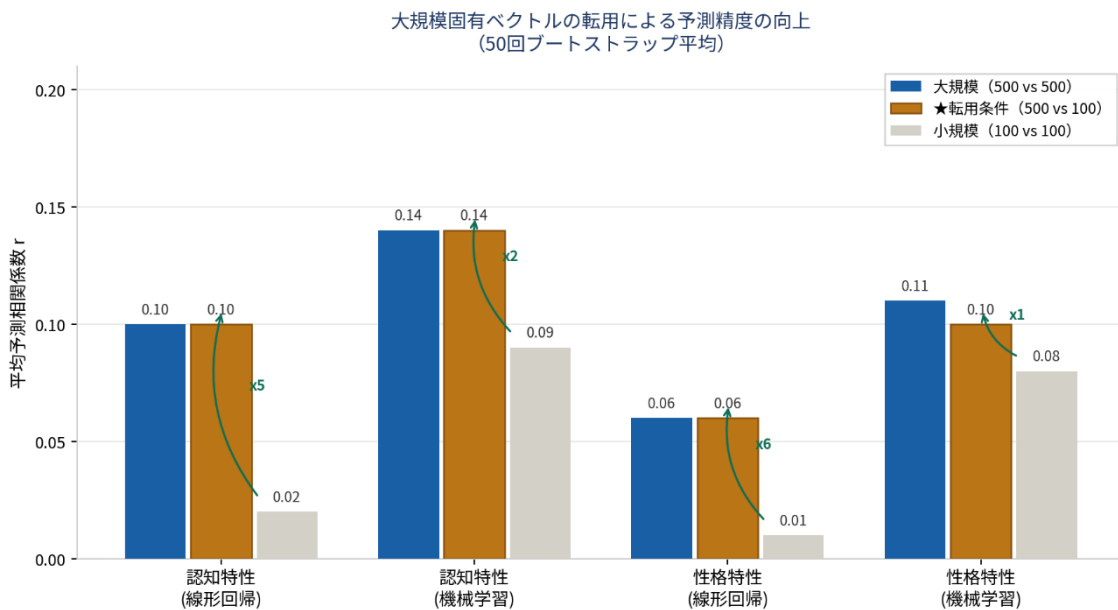


図 2. 新手法による予測精度の改善結果。100 名程度の小規模サンプルで、脳形態から行動を予測しても、精度は低い（灰）。しかし、大規模サンプルで抽出した脳形態の統計的特徴を小規模サンプルの解析に適用することで（赤）、大規模サンプル（青）並みの予測精度を得ることができた。

【謝辞】

本研究は、JST ムーンショット型研究開発事業 ゴール 9「2050 年までに、こころの安らぎや活力を増大することで、精神的に豊かで躍動的な社会を実現」（研究開発プロジェクト番号：JPMJMS229C）の支援を受けて実施されました。

【用語説明】

- 注1. 主成分分析（PCA）高次元データの次元を削減し、データの分散を最大化する方向（主成分）を求める統計手法。脳画像では 16 万 4 千次元のデータを圧縮するために用いる。次元の削減とは、たくさんの情報を、特徴を失わないように少数の指標にまとめること。
- 注2. 主成分ベクトル（固有ベクトル）PCA で算出される「主成分の方向」を定義するベクトル。大規模データから算出したものは安定しており、他のデータへの転用が可能。
- 注3. HCP（Human Connectome Project）ワシントン大学等が主導する脳神経回路地図作成プロジェクト。22～35 歳の健常成人 1,206 名の高解像度脳画像・行動データを無償公開。
- 注4. HDLSS 問題 高次元低サンプル数（High Dimension Low Sample Size）問題。変数数がサンプル数を大きく上回る状況で統計的推定が不安定になる問題。
- 注5. コサイン類似度 2 つのベクトルの方向の類似性を 0～1 の値で表す指標。PCA 成分の安定性評価に用いた。
- 注6. ハンガリアン法 複数の成分間で最適な一対一対応を見つける組合せ最適化アルゴリズム

【論文情報】

タイトル：Large-sample PCA eigenvectors stabilize cortical thickness components and improve small sample brain behavior prediction

著者：Zhang Yun Feng, Kenchi Hosokawa, Chihiro Hosoda

*責任著者：東北大学大学院情報科学研究科・加齢医学研究所 准教授 細田千尋（ほそだちひろ）

掲載誌：Scientific Reports

DOI : <https://doi.org/10.1038/s41598-026-52800-4>

URL : <https://www.nature.com/articles/s41598-026-52800-4>

【問い合わせ先】

(研究に関すること)

東北大学大学院情報科学研究科

准教授 細田 千尋

TEL: 022-795-5813

Email: chihiro.hosoda.d8@tohoku.ac.jp

(報道に関すること)

東北大学大学院情報科学研究科

広報室 小野寺 祐美子

TEL : 022-795-4529

Email: koho_is@grp.tohoku.ac.jp