



## PRESS RELEASE

岡山大学記者クラブ

文部科学記者会

科学記者会

御中

令和4年4月22日

岡山大学

慶應義塾大学

立教大学

### 脳活動を解読するブラックボックスの動作を深層学習で説明する新手法

#### ◆発表のポイント

- ・ 深層神経回路は高い精度で脳活動から行動を解読出来ますが、その動作の説明は困難です。
- ・ 本研究では、深層神経回路による脳活動解読を直感的に説明する新しい手法を提案しました。
- ・ 提案手法を応用して、深層神経回路が解読に使う脳活動の特徴を抽出することに成功しました。

岡山大学学術研究院自然科学学域（理・生物）の松井鉄平准教授と立教大学大学院人工知能研究科の瀧雅人准教授、生理学研究所のトラン・ファム(Trung Quang Pham)特任助教、株式会社アラヤの近添淳一主任研究員、慶應義塾大学（理工学部）の地村弘二准教授の共同研究グループは、脳活動を解読する深層神経回路の動作を直感的に説明する新しい手法を開発しました。

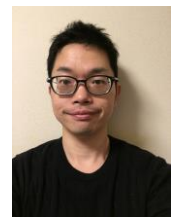
これらの研究成果は3月16日、スイスの神経科学雑誌「*Frontiers in Neuroinformatics*」の Research Article として掲載されました。

MRI や脳波により計測した脳活動データから、その人が何をやっていたのかを推定する脳活動解読は、BMI への応用を見据えて広く研究されている技術です。最近では、脳活動解読に深層学習を用いた研究が活発に行われています。しかし、深層神経回路によるデータの処理は非常に複雑で、「与えられたデータに対して解読器が何故その回答をするのか」を直感的に説明することが困難です。今回の研究では、この問題に対する新しいアプローチとして、深層生成モデルという深層学習のもう一つの技術と反実仮想説明という手法を組み合わせた手法を提案しました。

本研究成果は、認知症や精神神経疾患の脳画像診断に深層学習を応用していく際に、医師や患者が AI の動作を理解しつつ使用するための基礎技術として役立つと期待されます。

#### ◆研究者からのひとこと

ずっと人間や動物の脳を研究してきた神経科学者として、最近の深層学習の爆発的な発展には、とても刺激を受けています。今回は深層学習の技術を神経科学に応用してみましたが、生物の脳と深層学習を比較することで知性の本質を明らかにすることにもチャレンジしていきたいと思っています。



松井准教授



## PRESS RELEASE

### ■発表内容

#### <現状>

MRI や脳波により計測した脳活動データから、その人が何をやっていたのかを推定する脳活動解読は、BMI<sup>(1)</sup>への応用をはじめ広く研究されている技術です。最近では、脳活動解読に深層学習<sup>(2)</sup>を用いた研究が活発に行われ高い精度が報告されています。深層神経回路はコンピューター上で動くものなので、入力したデータが解読結果に至るまでの変換過程は全て記録されています。しかし、沢山の素子が複雑に絡み合っている深層神経回路の動作は非常に複雑で、「与えられたデータに対して解読器が何故その回答をするのか」という疑問に答えるのは難しいことが知られています。このことは、例えば解読器が間違った答えを返してきたときに、何故そのような間違いが起きたのかを説明することが出来ないということを意味しています。こうした欠点を克服し、複雑な深層神経回路の動作を説明可能にする技術の開発は、脳活動解読のみならず広く人工知能分野での重要なトピックとしても近年着目されています。

#### <研究成果の内容>

岡山大学学術研究院自然科学学域（理・生物）の松井鉄平准教授と立教大学大学院人工知能研究科の瀧雅人准教授、生理学研究所のトラン・ファム(Trung Quang Pham)特任助教、株式会社アラヤの近添淳一主任研究員、慶應義塾大学（理工学部）の地村弘二准教授の共同研究グループは、脳活動を解読する深層神経回路の複雑な動作を説明する新しい手法を開発しました。

今回の研究では、脳活動解読器の動作を直感的に説明するアプローチとして、反実仮想説明という手法を選びました。反実仮想説明では、実際のデータと少しだけ異なる仮想的なデータを作成して解読器に適用し、「もしデータが（実際とは異なり）X のようであったら、解読器は（実際に出てきた判断 Y と異なり）Z という判断をしていた」という分析を通して、直感的な説明を行います。この手法を脳活動解読器の説明に応用する上での難しい点は、脳活動のような複雑なデータに対して本物とそっくりな仮想的データを作るという部分でした。私たちは敵対的学習による深層生成モデルという深層学習のもう一つのテクニックを使うことで、本物そっくりな脳活動データを作ること成功し、反実仮想説明を脳解読器に適用できることを実証しました。図1では、脳活動解読器の誤作動を説明した例を示しています。本提案手法では、実際の脳活動データと、それを元に生成した仮想的脳活動データを比較することで、解読器の誤作動の原因になった脳部位を絵として表示し、直感的な説明を与えることができます。

また、この生成器の面白い応用として、特定の認知課題に向けて脳活動を「誇張」することで、その認知課題を特徴づけるパターンを取り出すことに成功しました。このパターンを、別の認知課題をしている時の脳活動に貼り付けて再び脳活動解読を行うと、解読の結果は元の脳活動ではなく貼り付けたパターンに対応するものとなり、脳活動生成器によって取り出したパターンが、実際に解読器によって使われていることが示されています。こういったパターン抽出は、今回の提案手法に特有の応用で、解読器の動作を基に新しい科学的な仮説を立てることなどに応用できると考えています。

PRESS RELEASE

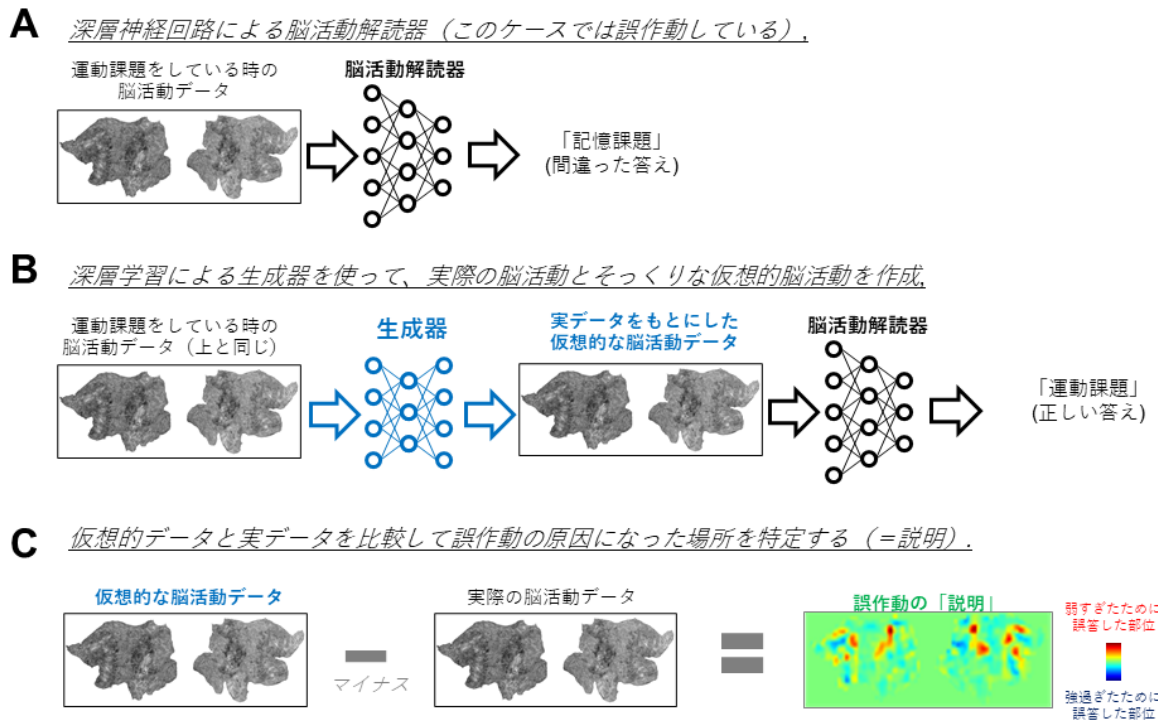


図 1. 提案手法によって脳活動解読器の動作の説明を与えた例。A) 運動課題を遂行中の脳活動データに対して、深層神経回路の解読器は「記憶課題」と間違えた答えを出した。複雑な解読器が何故誤作動したのかを説明することは、このままでは難しい。B) 解読器が間違えた脳活動データを生成器に入力し、それを元に仮想的な脳活動データを生成する。解読器はこの仮想的脳活動データを「運動課題」と正しく判断する。C) 仮想的脳活動データと実際の脳活動データを比較することにより解読器の「説明」が得られる。赤い脳部位が弱かったことや、青い脳部位がより強かったことが、解読器が回答を間違えた理由と考えられる。

<社会的な意義>

本研究成果は、認知症や精神神経疾患の脳画像診断に深層学習を応用していく際に、医師や患者が AI の動作を理解しつつ使用するための基礎技術として役立つと期待されます。

■論文情報

論文名 : Counterfactual Explanation of Brain Activity Classifiers Using Image-To-Image Transfer by Generative Adversarial Network

掲載紙 : *Frontiers in Neuroinformatics*

著者 : Teppei Matsui, Masato Taki, Trung Quang Pham, Junichi Chikazoe, Koji Jimura

DOI : <https://doi.org/10.3389/fninf.2021.802938>

URL : <https://www.frontiersin.org/articles/10.3389/fninf.2021.802938/full>



## PRESS RELEASE

### ■研究資金

本研究は、文部科学省科研費新学術領域研究（「脳情報動態を規定する多領域連関と並列処理」、課題番号 20H05052）、日本医療研究開発機構 AMED（「戦略的国際脳科学研究推進プログラム先進的個別研究開発課題」、課題番号 JP20dm0307031；「革新的技術による脳機能ネットワークの全容解明プロジェクト」、課題番号 JP21dm0207086）、JST さきがけ（「革新的コンピューティング技術の開拓」、課題番号 19205833）、文部科学省科研費学術変革領域B（「脳神経マルチセルラバイオコンピューティング」、課題番号 21H0516513；「情動情報解読による人文系学問の再構築」、課題番号 21H05060）の支援を受けて実施しました。

### ■補足・用語説明

1. BMI（Brain Machine Interface;ブレインマシンインターフェース）  
脳活動を計測して、それをもとに義手や義足などを動かす技術。
2. 深層学習（Deep Learning）  
元々は生物の脳にヒントを得て開発されたが、その後さまざまな工夫が加えられて現在の人工知能の基盤技術になっている。

#### <お問い合わせ>

岡山大学学術研究院自然科学学域（理・生物）  
准教授 松井 鉄平  
（電話番号）086-251-7860  
（メール）tematsui@okayama-u.ac.jp

