

ヒト脳内における視覚・意味情報の階層的処理の解明へ向けた取り組み

川崎 春佳 (指導教員：小林 一郎)

1 はじめに

近年、深層学習を使用してヒト脳の情報処理機構を解明する多くの研究が行われている。畳み込みニューラルネットワーク (CNN) を用いた研究では、脳内での視覚情報の階層的な処理が確認されている [1, 2]。また、言語モデルを用いた研究では、脳内での意味情報表現をモデル化可能であることが確認されている [3]。しかし、これら2つのモダリティについては個別に研究が進められることが多く、類似性や差異については十分に言及されてきていない。

本研究では、視覚情報および言語情報を扱う深層学習モデルを用いて符号化モデリング [4] を行い、予測を深層学習のモダリティと階層性から分析することにより大脳皮質上の視覚・意味情報の局在と表現内容の階層的特性を調査することを目的とする。

2 研究概要

深層学習モデルを使用して脳活動を予測するために、線形回帰を用いて計測脳活動パターンと予測脳活動パターンが近づくように重みを学習させる、符号化モデルを構築する。そして、大脳皮質上の視覚・意味情報の局在と表現内容の階層的特性を調査するために、符号化モデルの予測を使用した Representational Similarity Analysis (RSA) [5] を行う。RSA では、各ペアの非類似度を検索できる行列である Representational Dissimilarity Matrix (RDM) を作成して比較をする。

3 実験

3.1 実験設定

使用データ 使用した脳活動データは、情報通信研究機構脳情報通信融合センターにて機能的磁気共鳴画像法 (fMRI) により取得されたものである。

fMRI 実験には被験者7名が参加し、各被験者は2時間40分の音声付き自然動画を、固視条件 (スクリーン中央の固視点を注視する条件) と自由視条件 (自由に視線を動かす条件) のそれぞれで試聴した。

また、動画シーンの1秒ごとのキャプチャ画像に対して5.6名のfMRI実験に参加していないアノテーターから文章によるシーン記述を取得した。

特徴量の抽出 視覚情報として画像特徴量、意味情報として言語特徴量を用いた。

画像特徴量は VGG16 [6] の中間層8層から抽出した。動画の全フレームの画像特徴量を算出し、フレームレートの枚数分ずつで配列の要素ごとに最大値をとったものを、1秒分の画像特徴量として使用した。

言語特徴量は、事前学習済みの BERT_{BASE}¹ の中間層12層から抽出した。同じシーンを記述した全員分の特徴量を平均したものを1つのシーンに対応する特徴量として用いた。

符号化モデルの構築 画像特徴量に基づく符号化モデルと言語特徴量に基づく符号化モデルは同じ方法で作成した。固視および自由視条件で VGG16 (合計8層) と BERT (合計12層) の各階層から抽出した特徴量を使用し、合計40個の符号化モデルを構築した。特徴量の時系列を説明変数として脳活動の時系列を予測するモデルをリッジ回帰により学習した²。その際、神経活動に伴う血流の増加の反応時間を考慮し、fMRI で観測された脳活動データとその時系列の3, 4, 5, 6秒前の特徴量と回帰を行なった。また、チャンクを50として訓練データをシャッフルした上で10分割交差検証を行い、平均の相関係数が最も良くなる正則化項を採用した。

こうして学習したモデルを用いて、評価データの特徴量から予測した脳活動と同じ評価データが誘発し計測された脳活動でピアソンの積率相関係数を求め、それにより各ボクセルの予測精度を評価した。その際、False discovery rate で補正済みの p 値が有意 ($p < 0.05$) なボクセルを棄却した。

3.2 実験課題

大脳皮質上の視覚・意味情報の局在と表現内容の階層的特性を調査するために、上記の方法で構築した符号化モデルの予測精度と予測内容に対する RSA を行う。RDM 作成時には相関距離 ($1 - \text{ピアソン相関係数}$) を使用する。以下の全ての実験で、40個の符号化モデルのうち少なくとも1つの符号化モデルにおいて棄却されたボクセルのデータを用いる。RDM 作成方法を図1に示す。

(1) 予測可能な脳領域の類似性分析

40個の符号化モデルについて予測可能な脳領域の類似性を求めるためにこの分析を行った。全符号化モデルの予測精度を用いて (符号化モデルの総数) \times (符号化モデルの総数) の RDM を作成し、全被験者で平均した。

(2) 情報局在パターンの類似性分析

VGG16 の高層と BERT について、情報局在パターンの類似性を見るためにこの分析を行った。予測脳活動を用いて、符号化モデルごとに (ボクセル数) \times (ボクセル数) の RDM を作成した。

(3) 情報表現内容の類似性分析

40個の符号化モデルにおいて情報表現内容の類似性を求めるためにこの分析を行った。まず、予測脳活動を用いて、符号化モデルごとに (時間) \times (時間) の RDM を作成した。そしてこの RDM 全てを用いて (符号化モデルの総数) \times (符号化モデルの総数) の RDM を作成し、全被験者で平均した。

3.3 実験結果

実験課題 (1)~(3) の結果を以下に示す。

¹<https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

²リッジ回帰は <https://github.com/alexhuth/ridge> で提供されているものを用いた。

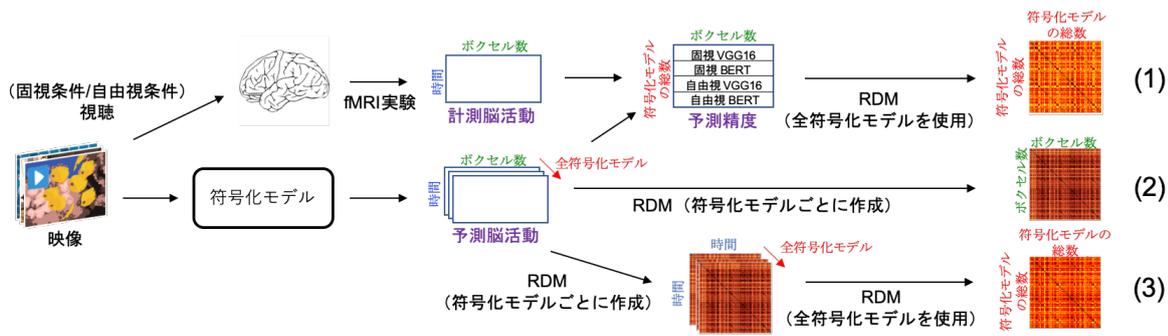


図 1: 予測精度および予測脳活動を対象にした RDM 作成 (RDM の右の項番は実験課題のものに相当する)

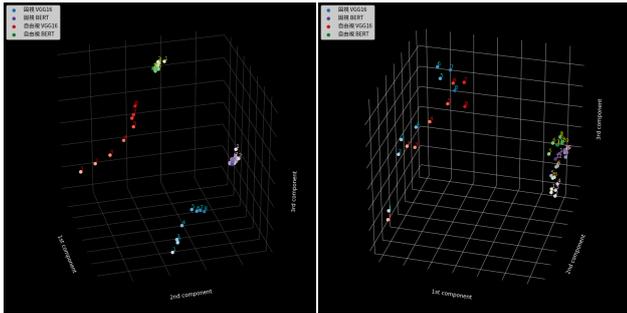


図 2: 予測可能な脳領域の類似性 図 3: 情報表現内容の類似性

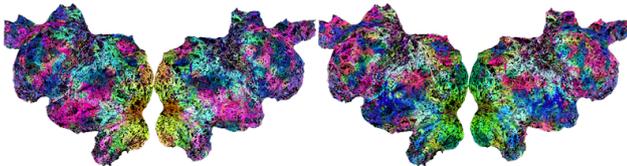


図 4: 自由視 VGG16 8 層 図 5: 自由視 BERT 12 層

(1) 予測可能な脳領域の類似性分析

可視化結果を図 2 に示す。これは (符号化モデルの総数) × (符号化モデルの総数) の RDM を MDS で (符号化モデルの総数) × 3 に圧縮し、3 次元空間上にプロットしたものである。近接して表示されるモデルほど似た脳領域を予測可能であることを示している。

(2) 情報局在パターンの類似性分析

1 名の被験者における自由視条件での VGG16 の 8 層・BERT の 12 層の可視化結果を図 4・図 5 に示す。これは RDM を UMAP で次元削減し、3 次元空間上で近いものを近い色、遠いものを離れた色となるように色付けした上で、その色を大脳皮質のフラットマップ上にプロットしたものである。それぞれの図で似た色のボクセル同士は似た情報表現を持ち、そうでないものは似ていない情報表現を持つことを示す。

(3) 情報表現内容の類似性分析

可視化結果を図 3 に示す。実験結果 (1) と同様の方法で次元削減、可視化を行った。近接して表示されるモデルほど似た情報表現内容を持っていることを示している。

3.4 考察

実験課題 (1) ではモデルの予測可能な脳領域について調査を行い、VGG16 の低層から高層に移るに従い予測可能な脳領域が BERT のそれと類似していくことが

分かった。また、実験課題 (2) より情報局在パターンが VGG16 の高層と BERT の高層で類似していることが視認された。これにより、VGG16 の高層と BERT の高層における情報表現内容は同じであると推定したが、実験課題 (3) のモデルの情報表現についての調査より、実際には VGG16 と BERT でモデル化できる情報表現内容には大きな違いがあることが分かった。つまり、同じ高次の感覚野においても、VGG16 と BERT は異なる脳内情報を表現していることが示唆された。

4 まとめ

本研究では、画像および言語に対して階層的処理を行う深層学習を利用し、中間層から脳活動を予測する符号化モデルを構築したうえで、予測精度と予測脳活動に対する RSA を通じて大脳皮質上の視覚・意味情報の局在と表現内容の階層的特性を調査した。

その結果、CNN と言語モデルは似た脳領域の情報局在をモデル化するが、モデルが捉えた情報表現内容は CNN と言語モデルで大きく異なることが分かった。

以上により、視覚情報は複雑になるにつれて意味情報と同じ脳領域で表現されるようになるが、その同一脳領域内においても視覚情報と意味情報の間には大きな表現の違いがあり、CNN と言語モデルによるモデル化だけではその違いを十分に埋められないことが示唆された。

参考文献

- [1] Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, and James J. Seibert, Darren and Di-Carlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, Vol. 111, No. 23, pp. 8619–8624, 2014.
- [2] Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, and Bertrand Thirion. Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, Vol. 152, pp. 184–194, May 2017.
- [3] Satoshi Nishida, Antoine Blanc, Naoya Maeda, Masataka Kado, and Shinji Nishimoto. Behavioral correlates of cortical semantic representations modeled by word vectors. *PLoS Computational Biology*, Vol. 17, No. 6, p. e1009138, 2021.
- [4] Thomas Naselaris, Kendrick N. Kay, Shinji Nishimoto, and Jack L. Gallant. Encoding and decoding in fmri. *NeuroImage*, Vol. 56, No. 2, pp. 400–410, May 2011.
- [5] Nikolaus Kriegeskorte, Marieke Mur, and Peter Bandettini. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, Vol. 2, p. 4, 2008.
- [6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, Vol. abs/1409.1556, , 2014.