

行列因子分解を用いた脳活動の言語表象推定への取り組み

川瀬千晶[†] 小林一郎[†] 西本伸志[‡] 西田知史[‡] 麻生英樹[§]
[†]お茶の水女子大学 [‡]情報通信研究機構 [§]産業技術総合研究所

1 はじめに

近年、脳神経科学分野において、脳神経活動を定量的に理解する研究が盛んに行われている [1]。本研究では動画視聴時のヒトの脳活動と、その動画を説明する文との対応関係をスパースコーディングにより分析し、初期視覚野におけるスパース表象と同様に高次表象である言語表象でも相同のスパースコーディングが行われているかを調査することを目的とする。これにより、情報のスパース性が脳皮質情報処理の一般原理となるかを調査する。

2 脳活動の言語表象推定

2.1 概要

図 1 に提案手法の概要を示す。

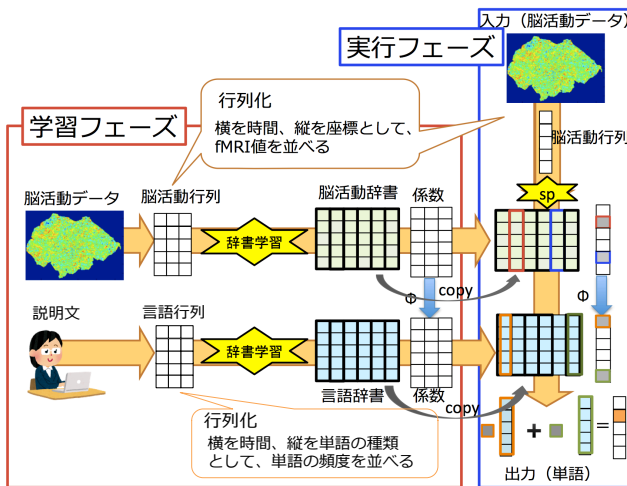


図 1: 提案手法の概要

本手法は、学習フェーズと実行フェーズに分けられる。学習フェーズでは、fMRI を用いて計測した脳活動データを時間ごとに計測した各部の観測値を入れて行列化し、これを脳活動行列とする。また、説明文も時間ごとに出現する単語頻度からなる行列を作り、これを言語行列とする。脳活動行列は辞書学習により、脳活動辞書行列と脳活動係数行列に分解する。同様に、言語行列も辞書学習により、言語辞書行列と言語係数行

列に分解する。次にリッジ回帰を用いて、脳活動係数行列を言語係数行列に写す写像 Φ を求める。実行フェーズでは、新たな脳活動データを入力として与え、脳活動行列を作成する。この行列を学習フェーズで作った脳活動辞書を用いてスパースコーディングをし、係数行列を求める。この係数行列を写像 Φ により、言語用の係数行列を求める。この係数行列と学習で作った言語辞書行列によって復元された行列の列 (2 秒) ごとに大きい値、上位 10 個に当たる単語を抽出し、脳活動に対応する言語表象の情報とみなす。

2.2 スパースコーディング

スパースコーディングとは、信号を少数の基底の線形和で表現する方法である。入力信号を辞書の中から、その入力信号の特徴となる少数の基底を選び、復元する手法である [2]。

$$X^* = \arg \min_X \frac{1}{2} \|Y - AX\|_2^2 + \lambda \|X\|_1 \quad (1)$$

X を係数、 Y を入力信号、 A を辞書とするとき、式 (1) を最小化する X を求めることにより、スパース (疎) な係数 X が得られる。argmin 内の第一項が入力信号 Y と復元信号 AX との二乗和誤差、第二項が係数のスパース性を表している。スパースコーディングにより入力信号を少数の基底の係数による線形和で表現可能となる。

2.3 辞書学習

$$(A, X)^* = \arg \min_{A, X} \frac{1}{2} \|Y - AX\|_2^2 + \lambda \|X\|_1 \quad (2)$$

ここでの辞書学習はスパースコーディングを用いた辞書学習であり、式 (2) により入力信号 Y を与えると係数 X がスパースになるような辞書 A と係数 X を求める。係数をスパースにすることにより密な辞書を作ることができる。

2.4 係数の回帰

本研究ではリッジ回帰を用い、脳活動データから得られた係数行列 X と説明文から得られた係数行列 Y 間における写像 Φ^* を求める。

$$\Phi^* = \arg \min_{\Phi} \frac{1}{2} \|Y - \Phi X\|_2^2 + \lambda \|\Phi\|_2^2 \quad (3)$$

式 (3) は最小二乗推定量に加え Φ の大きさに罰則をつけることで過学習を防ぐことを表している。

An approach to estimating Linguistic Representation of Brain Activity using Matrix Factorization

[†]Chiaki KAWASE(g1220516@is.ocha.ac.jp),

[†]Ichiro KOBAYASHI(koba@is.ocha.ac.jp)

[‡]Shinji NISHIMOTO(nishimoto@nict.go.jp),

[‡]Satoshi NISHIDA(s-nishida@nict.go.jp),

[§]Hideki ASOH(h.asoh@aist.go.jp)

3 実験

3.1 実験設定

使用するデータは、脳活動データと動画説明文である。脳活動データは、一人の被験者に動画画像を見せ、fMRIを用いてその時の脳神経活動を2秒で1サンプル記録したものである。脳活動の観測領域は $100 \times 100 \times 32$ ボクセルであり、今回の実験ではこの中からROI (Region of Interesting) 部分を対象とした8816ボクセルを使用した。各ボクセルの大きさは、 $2.24 \times 2.24 \times 4.1 \text{mm}^3$ である。動画説明文は被験者に見せた動画画像の2秒ごとの静止画の説明文を1つの静止画につき、アノテータ60人のうちランダムに抽出された5人が説明文を書いたものを使用した。このデータセットを訓練用データとして2394秒分、テスト用として510秒分それぞれを使用した。

3.2 評価方法

2秒ごとに、正解単語を自然言語説明文に出現する単語上位10個とし、実験で出力される10個の単語に対して評価を行った。脳活動データには動画視聴から約4秒程度の観測のずれがあるため、正解単語は、脳活動の観測ずれを考慮し、脳活動の4秒前の自然言語文に出現するものを抽出し、さらにそこから前後0秒、2秒、4秒の幅を広げたものとした。精度を推測された単語のうち正解単語に含まれるものの割合、再現率を正解単語のうち推測された正解単語の割合とし、動画ごとに精度、再現率、F値を求め、さらにその全体の平均値を求めた。

3.3 実験結果

訓練用データとして2394秒分の脳活動データを用い、学習させ、その後、テスト用データとして510秒分の脳活動データを動画ごとに分割し、それぞれの動画(35動画)に対して単語を推測し、全体をまとめて評価した(表1)。

表1: 実験結果の精度・再現率・F値

正解の幅	精度	再現率	F値
0秒	0.24	0.19	0.21
前後2秒	0.29	0.17	0.21
前後4秒	0.34	0.16	0.21

3.4 考察

表1より、精度、再現率、F値、共に低いことが分かる。推測された単語と正解単語について考察するために、再現率の最も高かったテストデータ1動画分(476~482秒目の動画)に対しての実験結果(表2)と正解の幅を0秒にしたときの正解単語(表3)を示す。

480秒目~486秒目*の脳活動データから推測された単語には、山、女性、人など同じ単語が上位に表れており、この動画を通して脳は似たような反応を示しているのではないかと考えられる。列ごとの単語同士を

*観測のずれを考量して脳活動データは正解データより4秒遅らせてある。

表2: 実験結果(1動画分)

480秒目	482秒目	484秒目	486秒目
山	山	山	女性
人	雪	女性	人
女性	波	人	顔
花	木	雲	着る
雪	れる	見る	何
夜	富士山	何	猿
後ろ	積もる	よう	アザラシ
信号	もの	れる	泳ぐ
パラ	光	犬	自転車
電車	山頂	明るい	サメ

表3: 正解単語(1動画分)

476秒目	478秒目	480秒目	482秒目
サメ	の	泳ぐ	泳ぐ
形	泳ぐ	魚	サメ
よう	サメ	よう	よう
頭	よう	人	姿
遭遇	イルカ	離れる	自由
変わる	何	目	部分
中	れる	の	大きい
泳ぐ	形	サメ	もの
海	こと	左右	の
写真	自分	たくさん	ヘッド

見てみると、例えば、482秒目には山、雪、富士山、積もる、山頂など関連のある単語が表れていることが分かる。言語辞書の基底を確認したところ、例えば5列目の基底の中で値の大きい単語上位10個は(パラ、花、花びら、綺麗、色、たくさん、山、咲く、店、きれい)であり、基底に特徴が表れていることが分かる。一方、3列目の基底は(犬、もの、タコ、白い、鳥、ちゃん、頭、横、怒る、こと)であり、これらの単語同士にはあまり関係がないように思われる。また、推測する単語は学習データに含まれる8175単語の中から抽出されるため、新しい単語を出力することはできない。本稿では、単語頻度行列をスパースコーディングを用いた辞書学習により言語辞書を求めたが、単語頻度行列は元々スパースな行列であるため、言語の情報を扱うには単語の意味を分散意味表現等で表した行列を対象にする等が考えられる。

4 まとめと今後の課題

本稿では、スパースコーディングを用いた辞書学習により、脳の活動部位から推測される言語表象を単語で表現した。しかし、脳の部位には、PPAやFFAのように視覚した物体のカテゴリーによって反応する領域があるため、語彙の類義性など意味的な広がりを捉える必要があると考えた。今後は言語データを分散意味表現に変えて、同様の実験を行う予定である。

参考文献

- [1] Tolga Cukur, Shinji Nishimoto, Alexander G Huth and Jack L Gallant "Attention during natural vision warps semantic representation across the human brain", 2013
- [2] Olshausen BA, Field DJ, "Sparse coding of sensory inputs.", Curr Opin Neurobiol, 2004.