

スパースコーディングを用いた脳活動の分散意味表現による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
[†]{g1220516,koba}@is.ocha.ac.jp, [‡]{nishimoto,s-nishida}@nict.go.jp,
[¶]h.asoh@aist.go.jp

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大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。本手法は、学習フェー

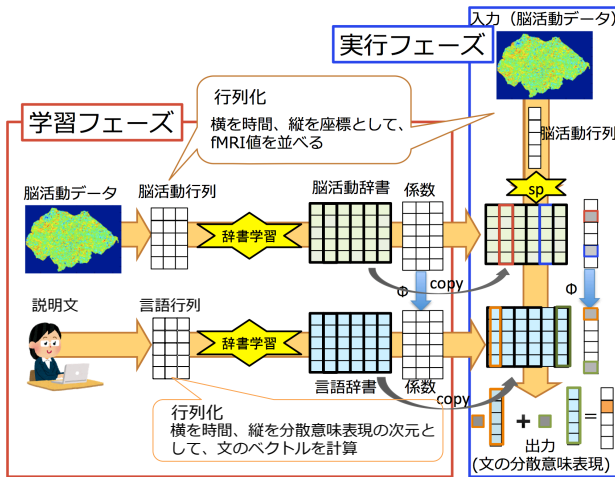


図 1: 提案手法の概要

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の分散意味表現の和のベクトル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を、脳活動に対応する言語表象の情報とみなす。

2.2 分散意味による説明文表現

言語行列を作る際、時間ごとに説明文に出てくる単語を、Mikolov によって提案された word2vec において、日本語 Wikipedia のコーパスを対象に Skip-gram を利用して構築した 1000 次元の分散意味表現を用いて表し、その和により文全体の分散意味とした。

2.3 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。

2.4 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。

2.5 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。

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。脳活動データには動画視聴から約 4 秒の観測のずれがあるため、言語データは、脳活動データの約 4 秒前のデータと対応するように学習した。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 540 秒分それぞれを使用した。また、本研究ではスパースコーディングのアルゴリズムには Lasso-LARS、辞書学習のアルゴリズムには LARS と Lasso-LARS を用いた。言語行列を作るときに使用する単語は名詞のみ、名詞と動詞、名詞と動詞と形容詞、名詞と形容詞の 4 通りについて実験した。

3.2 評価方法

2 秒ごとに、正解データを動画説明文の分散意味表現ベクトルとし、実験で出力される分散意味表現ベクトルに対して評価を行った。正解ベクトルと推測されたベクトルとの \cos 類似度を求め、マクロ平均をとり、 $[-1, 1]$ の値で評価した。

3.3 実験結果

言語データは、名詞と動詞、名詞と動詞と形容詞、名詞と形容詞、の 4 通りの単語の分散意味表現ベクトルの和として実験した。また、脳活動データと言語データのずれを 2 秒、4 秒、6 秒と変えて実験した。これらの場合について、それぞれ推測された分散意味表現ベクトルと正解データの分散意味表現ベクトルとの \cos 類似度のマクロ平均を求めた(表 1)。

表 1: 推定結果と正解ベクトルの \cos 類似度

品詞	2 秒	4 秒	6 秒
名詞	-0.0045	-0.0057	-0.0043
名詞, 動詞	-0.0050	-0.0061	-0.0054
名詞, 動詞, 形容詞	-0.0053	-0.0061	-0.0052
名詞, 形容詞	-0.0052	-0.0054	-0.0046

3.4 考察

どれも 0 に近い値になってしまい、似ているとも似ていないとも判断がつかなくなってしまった。また、辞書学習やスパースコーディングやリッジ回帰のパラメータ等、いろいろと変更して実験してみる必要がある。本研究では、文書に出現する単語の分散意味表現ベクトルの和をその文書の言語表象として実験したが、単語間の関係を捉えた言語表象にする必要があるのではないかと考えられる。

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分散意味表現により表現した。今回は言語データを word2vec を用いた単語の分散意味表現ベクトルの和としたが、今後は文書の分散意味を捉えた Glove[3] や paragraph2vec[4] などに変えて、同様の実験を行い、調査を続ける予定である。

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.
- [3] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean "Distributed Representations of Words and Phrases and their Compositionality", 2013.
- [4] Quoc Le, Tomas Mikolov, "Distributed Representations of Sentences and Documents", 2014.