

脳活動データからのスパースコーディングによる意味表象推定と基底の分析

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、動画像などを視聴した際の脳の活動パターンから人がどのような意味カテゴリを想起しているかを調査する研究が盛んになってきており、多くの新しい知見が得られている。Huthら [1] は、動画像中に現れる物体や動作を類義語体系である WordNet の語彙で表現し、動画像の刺激 (WordNet 語彙 [2]) と脳神経活動との関係について調査し、脳の皮質における意味のマップを作成した。Stansburyら [3] は、潜在的意味解析手法 LDA [4] を用いて、静止画に対して付与された語彙からシーンに対するラベル付けを教師なし学習で行い、その結果と静止画に対する脳神経活動の関係を結びつけ、カテゴリに対する脳の意味解釈の活動領域を明確にするとともにモデルを構築した。Cukurら [5] は、動画像中の物体に注意を払い認識する際に、どのように認識の意味形態が変化しているかを脳活動データから推定している。このように統計的な言語モデルは脳活動における感覚や文脈の情報に基づく表象表現を説明するのに適したモデルであることが指摘されてきたが、さらに近年、西本、西田らは、Mikolovら [6] によって提唱された word2vec を構築する際に採用された Skip-gram と呼ばれる言語モデルが潜在意味解析手法等のこれまでの言語モデルに較べて、より適していることを同じ実験設定の下で確認し、日本語 Wikipedia をコーパスとし、Skip-gram を利用することで得られる日本語の語彙の分散意味表現と血中酸素飽和度で計測される脳神経活動の間に相関関係が存在することを示している [7]。本研究では、脳の活動に対応する word2vec による表現を意味表象と呼ぶ。また、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意味表象でも、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検証する。

2 脳活動情報からの意味表象推定

脳活動データから意味表象を推定する際にスパースコーディングを用いる手法を提案する。図 1 に以下に説明する処理の概要を示しつつ、提案手法の概要を示す。まず、fMRI を用いて計測した脳活動データをサンプルごとに計測した各ボクセル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、その画像説明文もサンプルごとに出現する単語 (名詞、動詞、形容詞) の分散意味表現の和のベクトルからなる行列を作り、これを意味表象行列とする。これら 2 つの行列を縦に結合させ、脳活動と意味表象の結合行列を作成する。この結合行列に対し辞書学習を行い辞書と係数に分解する。これにより、辞書行列には、脳活動の特徴と意味表象の特徴が 1 列になった基底が作られ、係数行列は脳活動と意味表象において共通の基底になる。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辞書から脳活動部分のみを用いてスパースコーディングをし、係数を求める。この係数行列と学習で作成した辞書の意味表象部分の行列によって復元された分散意味表現ベクトルを脳活動に対応する意味表象とみなす。これをパラレルコーパスである動画の説明文の分散意味表現ベクトルと比較する。

3 実験

3.1 データ

使用するデータは、動画視聴時の脳活動データと動画説明文である [8]。このデータセットが訓練用に 3600

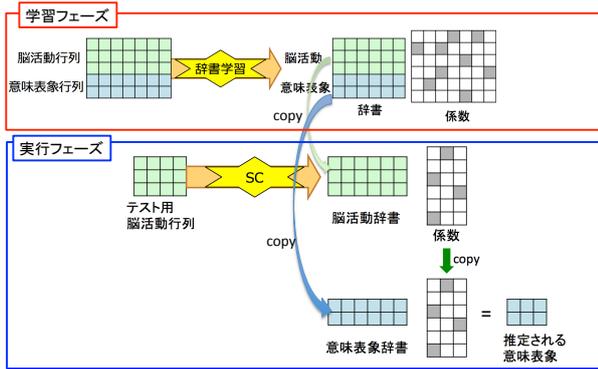


図 1: 提案手法の概要

サンプル, テスト用に 270 サンプルある. 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, そのうち大脳皮質部分が 30662 ボクセルある. 脳活動データの辞書学習をする際に, データ数 3600 サンプルよりもデータの次元を少なくしなくてはならないため, 30662 ボクセルのうち, 先行研究 [7] で予測精度が 0.36 以上の 1404 ボクセルを抽出した. 動画説明文は被験者に見せた動画画像から 1 秒ごとに抽出した静止画に対し, アノテータ 60 人のうちランダムに抽出された 5 人が静止画を見て書いた説明文を使用した. 説明文はその静止画を見て想起したことを書いてもらったものである. この脳活動データと画像説明文のデータを 2 秒分ずつ対応づける. また, 被験者が動画を見てから fMRI で観測される脳活動に影響が出るまでに約 4~6 秒の時間がかかる.

3.2 評価方法

テストデータの動画説明文に対しても学習データと同様に意味表象行列を作成し, これを正解行列とする. サンプルごとに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. また, スパースコーディングを用いずに脳活動行列と意味表象行列との回帰を学習することで意味表象を推定し, 提案手法の精度と比較する.

3.3 実験設定

脳活動行列の辞書学習アルゴリズムには Lasso-LARS と LARS を使い, 実行フェーズでのスパースコーディングには Lasso-LARS を用いた. 辞書の基底数は 2500 に設定した. また, fMRI の特性を考慮し,

脳活動データと言語データの時間を 2, 4, 6, 8 秒とずらして学習させ, 精度を比較した.

表 1: 推定精度

手法	辞書学習アルゴリズム	cos 類似度 / time lag			
		2 秒	4 秒	6 秒	8 秒
SC なし	-	0.066	0.176	0.190	0.120
提案手法	lasso-lars	0.142	0.244	0.247	0.222
提案手法	lars	0.139	0.247	0.250	0.181

表 1 から, スパースコーディングを用いない方法で意味表象を推定するよりも提案手法を用いて推定したほうが精度が向上した. 脳活動データと言語データのずれを 6 秒に設定した時の精度が最も高くなり, そこからずれが大きくなるほど精度が下がっていくことから, fMRI で観測される脳活動データに 4~6 秒の時間のずれが発生することと整合性がとれる. また, 最も精度の高かった辞書学習アルゴリズムを lars, データのずれを 6 秒にしたときの辞書を構成している基底について分析を行った.

3.4 基底の分析

辞書学習で作られた脳活動基底と意味表象基底の分布をそれぞれ可視化した. 図 2 は, これらの基底をそれぞれ主成分分析し, 寄与率の高い 3 次元を抽出し, その 3 次元空間上で表示した. 元の脳活動基底の次元数は 1404, 意味表象基底の次元数は 1000 である. 主成分分析の結果, 脳活動基底の主成分の寄与率は第一主成分が約 0.244, 第二主成分が約 0.068, 第三主成分が約 0.039 となり, これらの累積寄与率は約 0.350 となった. 意味表象基底の主成分の寄与率は第一主成分が約 0.095, 第二主成分が約 0.063, 第三主成分が約 0.041 となり, これらの累積寄与率は約 0.199 となった. 図を見ると偏りの少ない基底が生成されているように見える.

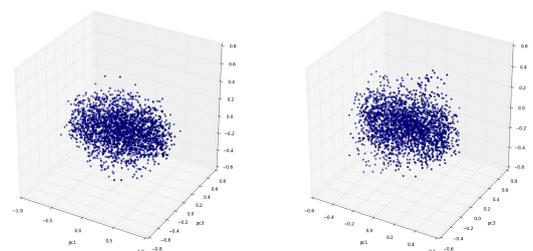


図 2: 基底の分布可視化 (3 次元)

そこで, 一方の基底を k-means 法によりクラスタ

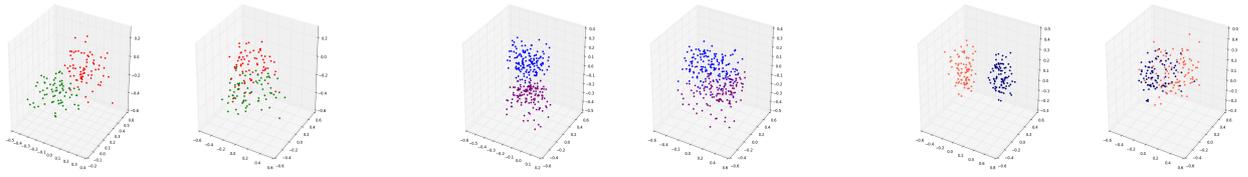


図 3: 脳活動基底のクラスタ (左) と対応する意味表象基底 (右)

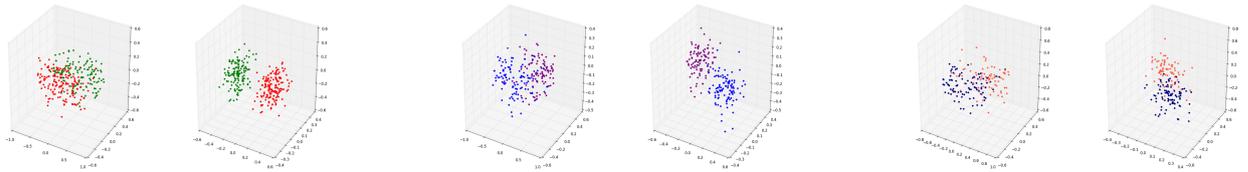


図 4: 意味表象基底のクラスタ (右) と対応する脳活動基底 (左)

数 $k(=30)$ を決め、クラスタリングを行いクラスタごとに色分けをし、それと対応するもう一方の基底に同じ色をつけ、それを主成分分析による 3 次元空間で表示した。k-means 法の距離には、脳活動基底にはユークリッド距離を、意味表象基底には \cos 類似度を用いた。脳活動基底 2500 個を 30 個のクラスタにクラスタリングし、それと対応する意味表象基底に同じ色をつけ、そのうちクラスタ 2 組ずつに含まれる基底の例を 3 つ表示した。左側に脳活動基底、右側に意味表象基底を示す (図 3)。多くの意味表象の基底は、脳活動のクラスタごとにまとまって現れた。また、意味表象基底を 30 個のクラスタにクラスタリングし、それと対応する脳活動基底に同じ色をつけ、そのうちクラスタ 2 組ずつに含まれる基底の例を 3 つ表示した。左側に脳活動基底、右側に意味表象基底を示す (図 4)。多くの脳活動の基底は、意味表象のクラスタごとにまとまって現れた。これらの図から、類似した脳活動基底と対応する意味表象基底は類似している場合が多いと考えられ、また、その逆も同様に考えられる。

4 おわりに

本研究では、ヒトの動画視聴時の脳活動データから意味表象を推定する際にスパースコーディングを用いる手法を提案し、スパースコーディングを用いないで推定するよりも精度が向上することを確認した。また、提案手法で得られた基底の分布を見ると、脳活動基底が類似していればそれと対応する意味表象基底も類似している場合が多いと考えられ、その逆も同様に考えられる。今後も辞書学習で生成された基底の分析を続

け、脳活動と意味表象の基底のもつ意味を理解する実験を行いたい。

参考文献

- [1] Huth AG, Nishimoto S, Vu AT, Gallant JL, "A continuous semantic space describes the representation of thousands of object and action categories across the human brain", *Neuron* 76(6):pp.1210-1224, 2012.
- [2] George A. Miller, "WordNet: A Lexical Database for English", *Communications of the ACM* Vol. 38, No. 11: pp.39-41, 1995.
- [3] Stansbury DE1, Naselaris T, Gallant JL, "Natural scene statistics account for the representation of scene categories in human visual cortex", *Neuron* 79(5):1025-34. *j.neuron*, 2013.
- [4] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research* 3, pp.993-1022, 2013.
- [5] Tolga Cukur, Nishimoto S, Alexander G Huth and Jack L Gallant, "Attention during natural vision warps semantic representation across the human brain", *Nature Neuroscience*, Volume 194, January 2013, pp.240-252, 2013.
- [6] T.Mikolov, I.Sutskever, K.Chen, G.Corrado and J.Dean, "Distributed Representations of Words and Phrases and their Compositionality", *Advances in Neural Information Processing Systems* 26, pp. 3111-3119, 2013.
- [7] Nishida S, Huth AG, Gallant JL, Nishimoto S, "Word statistics in large-scale texts explain the human cortical semantic representation of objects, actions, and impressions", *Society for Neuroscience Annual Meeting*, 2015.
- [8] Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL, "Reconstructing visual experiences from brain activity evoked by natural movies", *Current Biology* 21(19):pp.1641-1646, 2011.