

脳活動におけるスパースコーディングによる意味表象推定

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。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、先行研究 [9] で得られた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本質的に機能している可能性があるという知見の正当性をさらに検証することを目的とし、意味表象を構成する文章の分散意味表現の計算方法を変更し 2 つの手法の適用を行う。手法は動画視聴時の脳活動データから意味表象を推定するものであり、その際、1 つ目の手法では意味表象にスパースコーディングを適用し、2 つ目の手法では脳活動データと意味表象の両方にスパースコーディングを適用する。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 先行研究

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]。

3 脳活動情報からの意味表象推定

大脳皮質の意味表象においてもスパースコーディングが機能しているかを検証する方法として以下に示す 2 つの手法を提案する。

3.1 手法 1：脳活動行列のから意味表象係数行列への回帰に基づく意味表象推定

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、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、説明文から意味表象行列を作成する。意味表象行列はスパースコーディングを用いた辞書学習により意味表象辞書行列と意味表象係数行列に分解する。次に Ridge 回帰を用いて、脳活動行列を意味表象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。この脳活動行列を写像 Φ により写し、意味表象係数行列を求める。この意味表象係数行列と学習で作成した意味表象辞書行列によって復元された分散意味表現のベクトルを、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ことにより、意味表象

にスパースコーディングが機能しているかを検証する (図 1)。

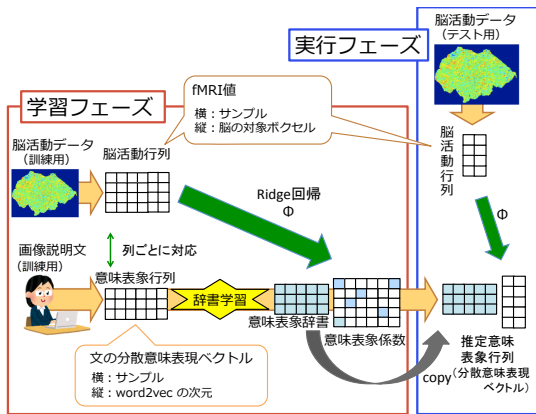


図 1: 手法 1 の概要

3.2 手法 2 : 脳活動係数行列から意味表象係数行列への回帰に基づく意味表象推定

次に、脳活動行列に対してもスパースコーディングを適用することにより、脳活動データの本質的な情報を抽出し、それに基づく意味表象を推定することを目指す。手法 1 と同様に、脳活動データから脳活動行列を作成し、説明文から意味表象行列を作成する。それぞれ脳活動行列と意味表象行列に対し辞書学習を行い、辞書と係数に分解する。次に Ridge 回帰を用いて、脳活動行列を意味表象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、脳活動係数行列を求める。この脳活動係数行列を写像 Φ により写し、意味表象係数行列を求める。この係数行列と学習で作成した意味表象辞書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 (図 2)。

3.3 意味表象の作成

意味表象行列を作る際、サンプルごとのアノテーションに出てくる単語 (名詞、動詞、形容詞) を、Mikolov[6] 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、文書に出現する単語の分散意味表現に、tfidf による単語重要度の重みをかけたものを文の意味表象ベクトルとした。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

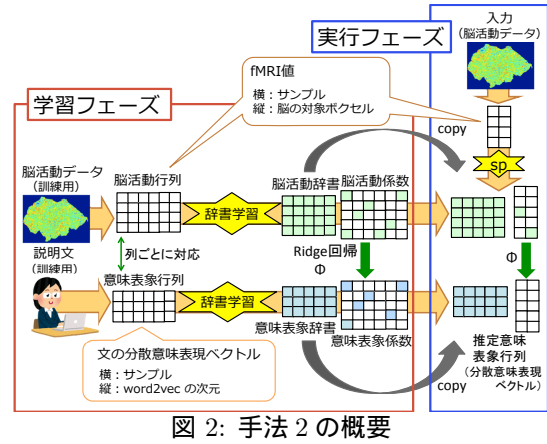


図 2: 手法 2 の概要

$n_{i,j}$ は単語 t_i のサンプル s_j における出現回数, $\sum_k n_{k,j}$ はサンプル s_j における全ての単語の出現回数の和とし、そのサンプルに多く出現する単語の重要度を上げる。

$$idf_i = \log \frac{|S|}{|\{s : s \ni t_i\}|} \quad (2)$$

idf_i を単語重要度で、 $|S|$ は総サンプル数、 $|\{s : s \ni t_i\}|$ は単語 t_i を含むサンプル数とし、多くのサンプルに出現する単語は重要度は下がり、特定のサンプルにしか出現しない単語の重要度をあげる (式 (2))。

$$tfidf = tf_{i,j} \cdot idf_i \quad (3)$$

式 (3) を計算し、この値をサンプルの特徴を表す単語の重要度とみなす。

4 実験

4.1 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 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、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、対象ボクセルとした。動画説明文は被験者に見せた動画像から 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。

4.2 実験設定

アルゴリズムは先行研究 [9] で最も良い結果が得られた組み合わせを用いた。すなわち、脳活動行列の辞書学習には Lasso-LARS と LAS を組み合わせ、意味表象行列の辞書学習には Lasso-LARS、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。基底数は意味表象辞書行列では 1200 または 1500、脳活動行列では 1500 に設定した。

4.3 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。

4.4 実験結果

表 1: 推定行列と正解行列との \cos 類似度

手法	意味表象辞書の基底数	\cos 類似度
Ridge 回帰	-	0.078
提案手法 1	1200	0.080
提案手法 1	1500	0.080
提案手法 2	1200	0.128
提案手法 2	1500	0.128

推定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示す (表 1)。また、提案手法との比較としてスパースコーディングを用いず脳活動行列から意味表象行列に直接 Ridge 回帰をして推定を行った。

4.5 考察

辞書の基底数を変化させて実験してみたが、提案手法 1 では Ridge 回帰のみによる推定より精度が向上し、提案手法 2 ではさらに向上した。このことから、意味表象においてもスパース性が機能しているとは断定はできないが、脳活動の特徴と対応する意味表象の特徴をスパースコーディングによって上手く捉えられたと考えられる。tfidf による単語重要度を導入することにより、一般語の影響がより減ったこともあり、推定がより難しくなったと考えられる。さらに \cos 類似度の高いサンプルと低いサンプルの差が激しくなり、刺激画像によって推定しやすいものと難しいものがあると考えられる。

5 まとめと今後の課題

本研究では、大脳皮質の意味表象においてもスパースコーディングが本質的に機能している可能性がある

という仮説の下、その正当性をさらに検証することを目的とする。先行研究 [9] で提案した手法 1 と 2 に対し、意味表象を単語の分散意味に単語重要度を考慮したものに变更し、動画視聴時の脳活動データからスパースコーディングを用いて意味表象の推定を行った。そして推定にスパースコーディングを用いることにより精度が向上したことを確認し、さらにこの仮説の正当性を高めた。今後の課題として、現在はアノテータから収集した全ての文書を用いて意味表象行列を作成しているが、実験に対するノイズとなる文も含まれているため、この中から適切な文だけを抽出して使用することも検討したい。

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.
- [9] Kawase C, Kobayashi I, Nishimoto S, Nishida S, Asoh H, "Toward Estimation of the Language Representation of the Brain Activity evoked by Visual Stimulation with Matrix Factorization", *The 30th Annual Conference of the Japanese Society for Artificial Intelligence*, 2016.