

スパースコーディングを用いた脳活動の意味表象推定に関する精度向上への取り組み

Toward Improving Estimation Accuracy of the Semantic Representation of Brain Activity Using Sparse Coding

川瀬千晶^{*1} 小林一郎^{*1} 西本伸志^{*2} 西田知史^{*2} 麻生英樹^{*3}
Chiaki Kawase Ichiro Kobayashi Shinji Nishimoto Satoshi Nishida Hideki Asoh

^{*1}お茶の水女子大学 Ochanomizu University ^{*2}情報通信研究機構 脳情報通信融合研究センター
National Institute of Information and Communications Technology

^{*3}産業技術総合研究所 人工知能研究センター
National Institute of Advanced Industrial and Technology

It is known that primary visual cortex uses a sparse code to efficiently represent natural scenes. Based on the fact, we build up a hypothesis that the same phenomenon happens at the higher cognitive function, here we focus on semantic representation, in the cerebral cortex. To proof the hypothesis, we estimated the semantic representation from the brain activity while s/he is watching video by using sparse coding. And we confirmed the improvement of accuracy for estimating the semantic representation by applying our proposed method employing sparse coding. Also, to confirm the validity of the method, we devise a way to make the semantic representation from corpus – only important annotated texts to the videos are used.

1. はじめに

ヒトの大脳皮質の初期視覚野での処理において、情報はスパースに処理をされていることが知られている [1]。初期視覚野には様々な刺激に反応する細胞が多数あり、視覚情報を受け取ると、その中から少数の細胞のみが反応し、入力信号を表現する。これにより、複雑な入力信号から本質的な情報を抽出し、効率よく視覚情報を処理している。また、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おり [2][3]、脳活動と言語表象の関係について様々な知見が得られている。本研究では、特に、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という仮説を立て、その仮説を立証することを目的とする。具体的な方法として、動画視聴時のヒトの脳活動の観測データと、その動画の説明文との対応関係がスパースコーディングを介在させることにより予測精度が向上するかを二つの手法を用いて確認する。

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 [3][4][5][6][7]。Huth ら [5] 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 [8]) 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 [9] 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 [2] 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 [11] によって提唱された word2vec を構築する際に採用された skip-gram 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 と呼ばれる言語モデル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. 脳活動情報からの言語表象推定

大脳皮質の言語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：脳活動行列からの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)。

連絡先: 川瀬千晶, お茶の水女子大学, g1220516@is.ocha.ac.jp

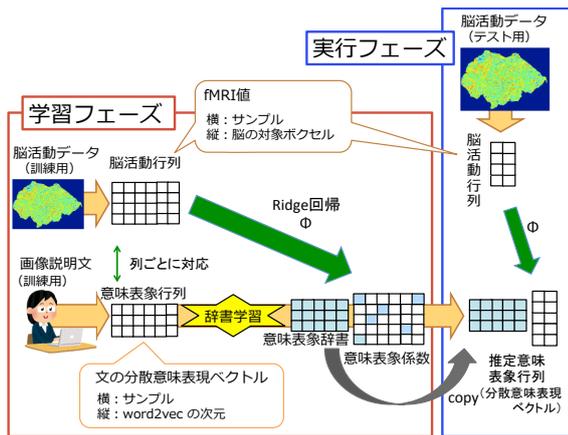


図 1: 手法 1 の概要

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)。

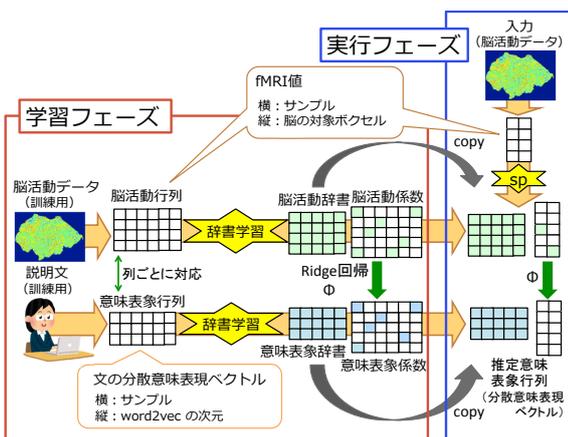


図 2: 手法 2 の概要

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[11]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の総和を文の意味表象ベクトルとした。アノテーションは 1 秒

で 5 人分の文書を含むため、1 サンプル(2 秒)には 10 人分の文書が含まれる。しかし、アノテーションの中には一般の人が想起しにくいような個人的な意見もあり、そのような文は今回の実験に使用するには適さない。そのため、それぞれのアノテータが書いた文書の重要度を以下のように計算し、重要度が閾値以上の文書のみを使用した。

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

$n_{i,j}$ 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。

$$score_{t_i} = \frac{\sum_i tfidf_{i,l}}{\sum_k n_{k,l}} \quad (4)$$

$\sum_i tfidf_{i,l}$ は文書 d_l に出現する全ての単語の重要度の総和であり、これを $\sum_k n_{k,l}$ (文書 d_l における全ての単語の出現回数の和)で割ることにより文長の影響を排除する。この値をサンプルの特徴を表す文書の重要度とみなし、 $score$ の値が閾値以上の文書 d_l のみに含まれる単語の分散意味ベクトルを足したものをそのサンプルの意味表象とする。

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[10]。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[15]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	-	0.080
提案手法 1	1200	0.200
提案手法 1	1500	0.200
提案手法 2	1200	0.102
提案手法 2	1500	0.102

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

意味表象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た。このことから大脳皮質での意味表象の処理にはスパースコーディングが機能している可能性があることが確認できた。提案手法 2 では今回の実験設定においては Ridge 回帰のみによる推定よりは精度が向上したが、提案手法 1 より精度が低くなった。このことはアノテータから得られた動画の説明文を全て使用した先行研究 [15] の結果とは一致しなかった。また、両手法とも基底に対する \cos 類似度の結果が同じであることから、今回の実験設定では脳活動の特徴と対応する意味表象の特徴が上手く捉えられていない可能性があり、スパースコーディングのアルゴリズムの組み合わせやパラメータの改善の余地がある。

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 [15] で提案した手法 1 と 2 に対し、意味表象に使用するアノテーションを重要度の高い文だけに絞り、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。そして意味表象をスパースコーディングを用いて処理す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。今後はスパースコーディングのアルゴリズムの組み合わせやパラメータの設定を見直し、再度実験を試みたい。

参考文献

- [1] Olshausen BA, Field DJ, "Sparse coding of sensory inputs", Current Opinion Neurobiology 2004, 14:481-487.
- [2] 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.
- [3] Stansbury DE1, Naselaris T, Gallant JL, "Natural scene statistics account for the representation of scene

categories in human visual cortex", Neuron79(5):1025-34. j.neuron,2013.

- [4] Francisco Pereira, Matthew Botvinicka, Greg Dettre, "Using Wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments", Volume 5, Article 72, 2011.
- [5] 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", Neuron76(6):pp.1210-1224,2012.
- [6] T. Horikawa, M. Tamaki, Y. Miyawaki, Y. Kamitani, "Neural Decoding of Visual Imagery During Sleep", SCIENCE VOL 340,2013.
- [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] George A. Miller, "WordNet: A Lexical Database for English", Communications of the ACM Vol. 38, No. 11: pp.39-41,1995.
- [9] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, pp.993-1022, 2013.
- [10] 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.
- [11] 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.
- [12] Jeffrey Pennington, Richard Socher and Christopher D. Manning, "Glove: Global Vectors for Word", Conference on Empirical Methods in Natural Language Processing (EMNLP 2014),2014.
- [13] Quoc Le, Tomas Mikolov, "Distributed Representations of Sentences and Documents", Google Inc, 1600 Amphitheatre Parkway, Mountain View, CA 94043,2014.
- [14] William E. Vinje and Jack L. Gallant, "Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision", Science 18, Vol. 287, Issue 5456, pp.1273-1276,2000.
- [15] 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.