The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

脳活動と分散表現による意味表象へのスパースコーディング適用により獲得された辞書基底の分析

Analysis of the Dictionary Bases obtained by Applying Sparse Coding to Brain Activity and the Semantic Representation based on Distributed Representation

川瀬千晶 *1 Chiaki Kawase　小林一郎 *1 Ichiro Kobayashi　西本伸志 *2 Shinji Nishimoto　西田知史 *2 Satoshi Nishida　麻生英樹 *3 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 applied sparse coding to the brain activity while the subject is watching the video and the semantic representation of the video. By means of this method, we obtained the dictionary matrix consisting of bases which represent the corresponding relation between the brain activity and the semantic representation, and we analyzed the characteristics of each basis.

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 [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 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，それぞれの基底がどの様な機能を表現しているかについて調査する．

2. 脳活動と意味表象の辞書学習

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（図 1）．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．このようにして作られた辞書を構成する基底について分析を行う．

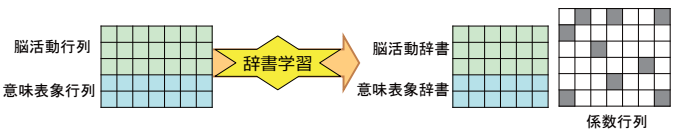


図 1: 脳活動データと意味表象の対応を保持する辞書学習

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 [10]．このデータセットを 3600 サンプルを使用する．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 100×100×32 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いけないため，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．

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

3.2 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 を用い，辞書の基底数は辞書学習を行う際のデータ数，データの次元数と基底数の制約条件を考慮し，2500 に設定した．

4. 基底の分布

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．

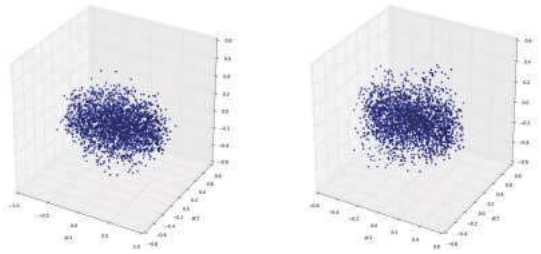


図 2: 基底の分布可視化（3 次元）

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，その逆も同様に言えると考えられる．

5. 基底の分析

5.1 実験方法

5.1.1 意味表象基底の表す意味

辞書学習により得た辞書の意味表象辞書の部分において各意味表象基底が表す意味を分析する．意味表象基底の持つ意味は単語で表せるものではないが，単語での近似を行うことで意味を理解する目安とする．各意味表象基底と本研究で word2vec 空間を学習するのに用いた word2vec コーパスに含まれる全単語との cos 類似度を計り最も小さい単語から上位 3 単語とその cos 類似度を出力し，これらの単語を各意味表象基底の表す意味の近似単語とした．

5.1.2 同じ基底を用いて復元されるサンプル同士の関連性

特定の基底を大きな重みで使って復元されるサンプル同士に関連が見られるかを検証する．これは，特定の基底を大きな重みで使って復元されるサンプル上位 5 つに対応する刺激画像同士に関連があるかを見ることで行う．基底には画像を見たときの脳活動と意味表象の特徴が表されると考えられるので，人が見て近い印象を抱く画像同士が出力されることが予想される．特定の基底に対応する係数ベクトルの非ゼロ要素の中で最も重みの大きいサンプル上位 5 つに対応する刺激画像とその重みを出力する．動画 1 サンプルは 2 秒間であり，その中央である 1 秒目の静止画像を出力した．

5.1.3 基底と ROI との関係

ヒトの大脳皮質には，ROI（Region Of Interest）という部位に分けられる部分がある．そのうち特定のカテゴリに反応する 2 つの ROI と基底との関係について分析を行った．1 つは FFA（fusiform face area）という顔に強い反応を示す領域，もう 1 つは PPA（parahippocampal place area）という風景や場所に強い反応を示す領域とした．本研究で実験対象とした脳活動データ 1404 ボクセルのうち FFA に属するボクセル数は 57 であり，PPA に属するボクセル数は 14 であった．これら 2 つの ROI が大きく寄与する基底に対して，近似単語と，その基底を用いて復元されるサンプルに対する刺激画像を出力し比較を行い，そのカテゴリに違いが見られるかを検証する．これらの ROI が大きく寄与している基底として，脳活動行列の対象 ROI に属しているボクセルに対する値を全て足した値が大きい基底上位 4 つを抽出し，これらの基底について実験を行った．

5.2 実験設定

基底は前から順に番号を付与する．まず，ランダムな基底の代わりに前から順に 4 つの基底に対して実験を行った（表 1）．次に，FFA が大きく寄与している基底上位 4 つ（基底 1585，基底 745，基底 563，基底 719）について実験を行った（表 2）．また，PPA が大きく寄与している基底上位 4 つ（基底 1436，基底 1364，基底 2400，基底 590）について実験を行った（表 3）．

5.3 実験結果

表 1，2，3 には，各基底番号，その基底の意味表象部分の近似単語上位 3 つとその cos 類似度，出力された刺激画像と画像の下にその重みを 1 位から 5 位まで左から順に示す．

5.4 考察

基底 0〜4 に対して実験を行ったところ，似たようなイメージの刺激画像同士が得られた．ここでは 4 例しか示していないが，各基底における実験結果についての考察を述べる．基底 0 においては，動物，特に動物の群れが映っている画像が多く出力された．基底 1 においては，1 位，2 位のサンプルの重みがそれ以下の重みと比べて高くなっており，これらの画像には男女が映っている．基底 2 においては，1 位，2 位，3 位の刺激画像は同じ動画内から抽出された画像となっており非常によく似ている．また，5 位の画像にもこれらの画像と類似した模様のある生物が映っている．基底 3 においては，1 位，3 位，5 位の画像には文字が見られた．2 位，4 位の画像には文字は映っていないが，その前後のサンプルと対応する画像を見てみたところ，文字が映っている画像であることを確認した．これは，動画 1 サンプルは 2 秒間であり，画像説明文はその中の

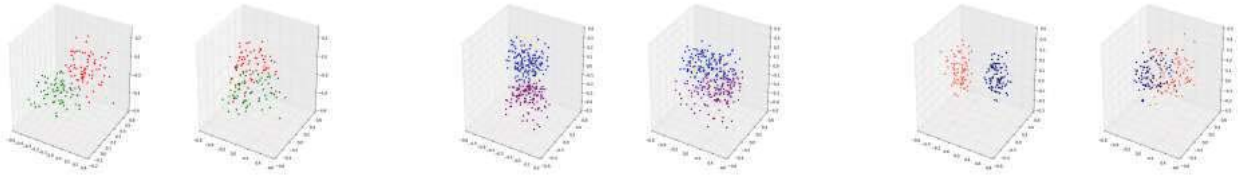


図 3: 脳活動基底のクラスタ（左）と対応する意味表象基底（右）

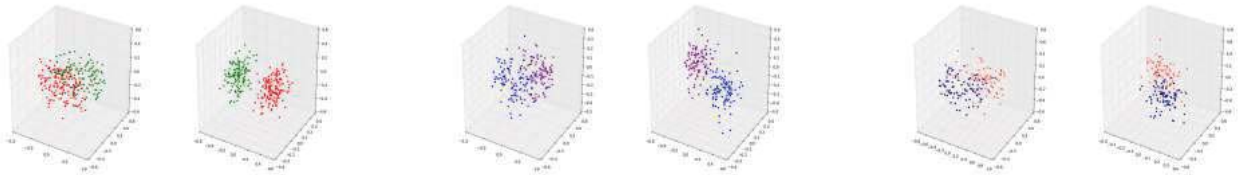


図 4: 意味表象基底のクラスタ（右）と対応する脳活動基底（左）

表 1: 特定基底 (ランダム) が大きく使われるサンプル

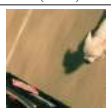
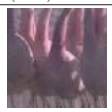
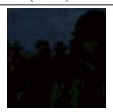
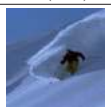
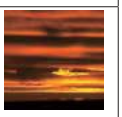
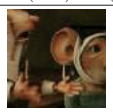
近似単語：和室 (0.23)，洋室 (0.22)，ミュージアムショップ (0.22)					
基底 0	 18.7	 15.3	 14.3	 12.8	 11.9
近似単語：粘板岩 (0.34)，吹き上げる (0.34)，等高線 (0.33)					
基底 1	 34.5	 24.4	 4.40	 3.57	 2.84
近似単語：ウツボ (0.60)，ウミヘビ (0.59)，アメフラシ (0.59)					
基底 2	 41.3	 17.1	 12.3	 6.31	 4.07
近似単語：文字 (0.40)，タイポグラフィ (0.35)，シンタックス (0.34)					
基底 3	 45.5	 7.45	 6.46	 4.77	 4.52

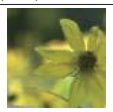
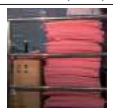
表 2: 特定基底 (FFA が強い) が大きく使われるサンプル

近似単語：男性 (0.36)，雇い主 (0.31)，チカチーロ (0.29)					
基底 1585	 46.7	 7.92	 7.12	 5.62	 3.90
近似単語：スーツ (0.41)，目元 (0.41)，いでたち (0.40)					
基底 745	 41.0	 12.8	 5.64	 5.29	 3.65
近似単語：パーティー (0.36)，女性 (0.35)，会食 (0.34)					
基底 563	 133	 109	 108	 96.8	 96.5
近似単語：包帯 (0.47)，服 (0.45)，ネズミ (0.43)					
基底 719	 49.4	 10.1	 5.98	5.40	3.20

0.5 秒目と 1.5 秒目の静止画像に対するものを合わせているのに対し，その中の 1 秒目の静止画像を出力したことが原因になっている可能性がある．このような例は他の基底においても見られたことから，刺激画像の抽出方法には改善の余地があると考える．他の基底について実験を行っても，同じ基底を大きな重みで使って復元されるサンプル同士は類似している例が多く見られた．このことから，辞書学習により意味表象の特徴を捉えた基底が得られ，これらの意味表象基底はさまざまな意味表象を表現するのに効率の良さに役立っていると考えられる．また，今回は辞書学習の条件の下，訓練データ数とデータの次元数を考慮した上で基底数を 2500 と設定したが，基底数を大きくするほど細かい特徴を捉える基底が得られやすく，基底数を小さくするほど大まかな特徴を捉える基底が得られ易くなると考えられ，最適な基底数を見つけることが重要だと考える．

また，意味表象基底の近似単語について考察を行うと，近似単語と刺激画像とに関連がある例（基底 2，3）とない例（基底 0，1）が見られた．これは，基底 2，3 においては近似単語との cos 類似度が高いのに対し，基底 0，1 においては低いことが関係していると考えられる．このことから，意味表象基底にはその意味を単語で近似しやすいものと近似しにくいものがあると考えられる．FFA と PPA についての実験結果からは，意味表象基底の近似単語を見ると，比較的に FFA の寄与が大きい基底は人を連想させるような単語が多く，PPA の寄与が大きい基底は風景や場所を連想させるような単語が多いことを観測した．出力された刺激画像においても FFA と PPA のカテゴリの違いが多少見られたものの，特に FFA の寄与率が 3 番目に大きい基底 563 に対して出力された刺激画像はどれも風景を表しており予想と反する結果となっており，この点に関

表 3: 特定基底 (PPA が強い) が大きく使われるサンプル

近似単語：地層 (0.54)，堆積岩 (0.52)，火山岩 (0.51)					
基底 1436	 48.6	 8.48	 7.61	 4.79	 4.22
近似単語：道路 (0.32)，街路 (0.30)，街並 (0.29)					
基底 1364	 48.3	 7.69	 3.82	 3.48	 3.47
近似単語：工場 (0.40)，アウトソーシング (0.31)，テクノロジーズ (0.31)					
基底 2400	 47.0	 18.6	 16.5	 14.1	 9.80
近似単語：並木道 (0.49)，小道 (0.46)，木漏れ日 (0.45)					
基底 590	 48.5	 8.50	 5.20	 3.74	 3.60

してはさらなる検証が必要である．

6. おわりに

本研究では，ヒトの動画視聴時の脳活動データと意味表象を辞書学習することにより，これらの特徴を捉えた基底を作ることに成功した．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．また，各意味表象基底の表現する意味を単語で近似した．また，同じ基底を大きな重みで使い復元されるサンプルには関連性が見られることを確認した．このことから，スパースコーディングにより得られた意味表象基底は，さまざまな意味表象を表現するのに効率の良いような特徴を捉えていると考えられる．さらに，2 種類の ROI の寄与が大きい基底での実験を行い比較をしたところ，これらのカテゴリに多少の違いが見られたがさらなる検証の余地が残されていることを観測した．今後はさらに実験を進め，脳活動と意味表象の対応関係についてより深い考察を進めたい．

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 Detre,“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. Jorfan,“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,“Semantic representation in the cerebral cortex with sparse coding”,2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017.