

3M-05

スパースコーディングを用いた脳活動の意味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

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、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

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。

2.1 手法1：脳活動行列の意味表象行列への直接回帰による意味表象推定

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。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)。

2.2 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

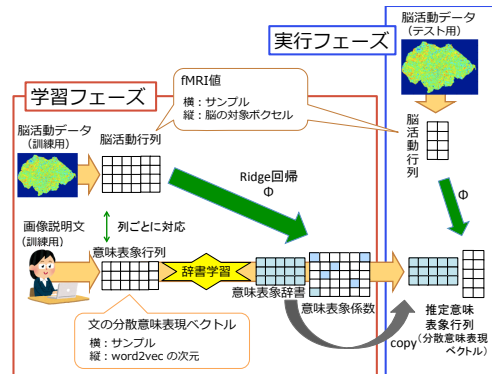


図1: 手法1の概要

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)。

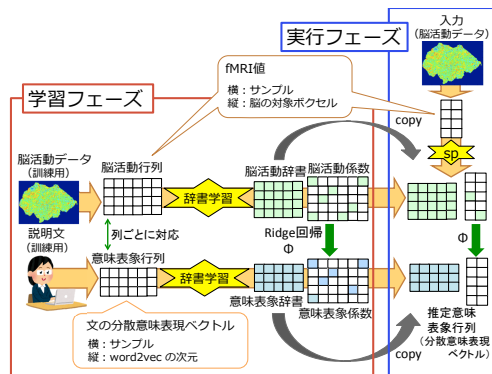


図2: 手法2の概要

3 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[1]。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100×100×32ボクセルであ

An Approach to Predicting the Semantics Representation of Brain Activity using Sparse Coding
[†]Chiaki KAWASE [‡]Ichiro KOBAYASHI [‡]Shinji NISHIMOTO
[‡]Satoshi NISHIDA [¶]Hideki ASOH

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 [2] 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．

3.2 動画切り替え時のサンプル除去

脳活動データは 2 秒で 1 サンプルのデータであるため，それと対応するアノテーションデータも 2 秒で 1 サンプルとする．意味表象ベクトルを作る際，2 秒分の文書を 1 サンプルにするため，前の動画の文書と次の動画の文書が足されたものができてしまい，そのとき違う動画に対する内容の文書が 1 つになった意味表象ベクトルができてしまう．そのため，動画の切り替え部分のサンプルの除去を行った上で実験をし，スパースコーディングを介して推定したほうが精度が上がるのかを検証する．動画の切り替えを抽出するために，アノテーションデータから名詞，動詞，形容詞を抽出する．そして文書に出現するそれぞれの単語に対して $tfidf$ (式 (3)) を計算し，サンプル毎にそれらの単語の $tfidf$ のベクトルを作る．そして，隣り合ったサンプル同士の \cos 類似度を計算し，閾値以上ならそのサンプル間で動画の切り替えが起こったとみなしそのサンプルを削除する．また，それに対応する脳活動データも削除する．

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

$n_{i,j}$ 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 (式 (1)) ．

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

$|S|$ 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) をサンプルの特徴を表す単語の重要度とする．

3.3 実験設定

今回，動画切り替えの閾値は 0.05 とした．アルゴリズムは切替部分除去を行わず推定を行ったときに一番精度の高かった組み合わせで実験を行った．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 または 2000 に設定した．

3.4 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．

3.5 実験結果

表 1: 推定行列と正解行列との \cos 類似度

手法	脳活動辞書の基底数	意味表象辞書の基底数	\cos 類似度
Ridge 回帰	-	-	0.150
提案手法 1	-	1200	0.161
提案手法 1	-	1500	0.158
提案手法 2	1500	1200	0.273
提案手法 2	1500	1500	0.271
提案手法 2	2000	1200	0.273
提案手法 2	2000	1500	0.271

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．

3.6 考察

動画の切替部分の判定は上手くいっているようであり，その部分のデータを除去すると訓練データ数は 3600 サンプルから 2590 サンプルに減った．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 ではさらに大きく向上した．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まだ言い難いが，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．

4 まとめと今後の課題

今回，動画切り替え部分のサンプルを削除した上で，動画視聴時の脳活動データから意味表象を推定する実験を行い，スパースコーディングによる精度向上が見られた．また，意味表象を作る際，今は全てのアノテーションを使っているが，今後はこの中から適切な文だけを抽出して使用することを考えている．

参考文献

- [1] 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.
- [2] 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.