スパースコーディングを用いた脳内意味表象推定における BERT の有効性の検証

島 百子　(指導教員：小林一郎)

1　はじめに

近年，脳神経科学において自然言語処理技術を導入し脳活動の解析を行うアプローチが盛んになっている．Jat ら [1] は，BERT（Bidirectional Encoder Representations from Transformers）[2] により表現された文がその文を聴いた被験者の脳活動データ (MEG) と強い相関があることを示している．一方，Kawase ら [3] は，動画を視聴した際の脳活動データと，動画説明文を word2vec によって表現したベクトル（本研究では「単語表象行列」と呼ぶ）との対応関係をとった結合行列に対してスパースコーディングによる辞書学習を行い，その辞書を用いて脳活動データの行列から脳内の意味表象行列を推定した．その結果，直接脳活動行列から単語表象行列への Ridge 回帰を推定した意味表象行列に比べ精度が高かったことから，脳内意味表象におけるスパースコーディングの原理の成立を仮説している．本研究では，言語情報の表象を word2vec から BERT に変更したもの（本研究では「文表象行列」と呼ぶ) を利用し，スパースコーディングによる解析における word2vec と BERT の性能を比較，調査する．

2　意味表象推定実験

一般に，脳神経科学の分野では脳内に持つ意味の情報を総称して「意味表象」という用語を使用するが，本研究においては，脳活動データから推定される言語の情報を総称して「意味表象」と呼ぶ．とくに，word2vec によって表現される意味情報として「単語表象」，BERT によって表現される意味情報を「文表象」と呼ぶ．

2.1　推定方法

図 1 に意味表象推定方法についての概要を示す．

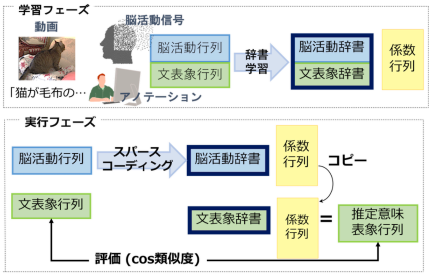

図 1: 意味表象推定方法の概要

2.2　脳活動データと言語データ

使用するデータは，VideoBlocks [5](以下 VB) および NishidaVimeo [6](以下 NV) データセットの動画視聴時の脳活動データと動画説明文である．VB データは動画の中心を注視するよう指示された被験者 A，B，C の 3 人分の脳活動データを保持し，訓練データが A と B は 4500 サンプル，C は 9000 サンプル，テストデータが A と B は 300 サンプル，C は 600 サンプルである．一方，NV データは視線を自由にした状態で計測した被験者 D，E，2 人分のデータを使用し、両者とも訓練データが 7200 サンプル，テストデータが 600 サンプルである。脳活動データは，fMRI(functional MRI) を用いて神経活動と相関があるとされる BOLD(Blood Oxygenation Level Dependent) 信号をボクセル数×サンプル数で記録したものであり，被験者 A と B については 2 秒に 1 サンプル，その他の被験者については 1 秒に 1 サンプル記録している。ただし，スパースコーディングを適用するにはボクセル数が膨大なため，二段階で次元削減を行った．まず，解剖学的な見地からの関心領域（ROI）に基づき，全脳から大脳皮質領域のみを取り出した．二段階目として、Nishida ら [6] は word2vec を用いた脳活動の推定モデルを構築し，ボクセルごとにピアソン相関係数を用いた推定精度を示していることから，閾値以上の推定精度を持つボクセルを抽出した．また，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像に対しアノテータが想起したことを文章にしたものである．アノテータは被験者とは別に用意した 40 人 (VB) または 48 人 (NV) で、このうちランダムに抽出された 4 人 (VB) または 4~7 人 (NV) の文章を合わせて動画 1 サンプルに対する説明文としている．ただし，NV データのテストデータについては，動画視聴後の被験者がアノテーションを作成し，2 人分の文章を 1 サンプルの動画説明文としている．

2.3　BERT を用いた辞書学習

まず、訓練用脳活動データと対応する言語データの結合行列を辞書学習し、両データが紐づいた辞書を作成する．学習には Lasso (Least Absolute Shrinkage and Selection Operator) の LARS アルゴリズムを用いる．脳活動データは BOLD 信号の値をサンプルごとに 1 列に並べ行列化した．このとき，先行研究 [4] では予測精度 0.55 以上のボクセルのみを利用しているが，本研究では言語データの分散表現が 300 次元から 768 次元に増えることを考慮し，脳活動データも同程度の次元数になるよう閾値を 0.48~0.57 に設定した．また，言語データについて川瀬らは word2vec を用いて出現単語を 300 次元の分散表現ベクトルで表現し、その平均を 1 サンプルのベクトルとしている．本研究では，言語情報の表象において言語学習モデル BERT（Bidirectional Encoder Representations from Transformers）を利用した．本モデルは双方向学習による文脈を捉えた特徴表現抽出を行っており，様々な言語学習タスクにおいて精度向上が報告されている．特に文単位で異なる意味空間を作るため，多義語に対応できるという特徴をもつ．京都大学黒橋・河原研究室が公開している BERT の Whole Word Masking 版日本語事前学習モデル (12-layer, 768-hidden, 12-heads) を用いて，アノテーションデータ 1 サンプル分を 1 シーケンスとして学習し，抽出した 768 次元のベクトルをサンプル数分並べ行列化した．図 2 に言語データの表象方法について先行研