

ヒト脳における時間認識時の脳内状態の推定

理学専攻・情報科学コース

2140660

須藤 百香

1 はじめに

近年、脳神経活動の多点計測技術の発展や深層学習に代表される機械学習技術の高度化により、観測したヒト脳内情報に対する解析や定量的理解を行う研究が盛んになっている。特に、汎用言語モデルと脳活動データの対応関係をとることによりヒト脳内の様々な特徴を解析することが可能であると証明した先行研究が多く存在する。このような背景を踏まえて、本研究では自然動画を視聴している際のヒト脳内情報と動画像内の発話を書き起こしたものから時制に関する部分を取り出し、符号化モデルを構築し、発話刺激によるヒト脳内の活動状態を推定することを試みる。

2 時間認識時の脳活動状態推定

本研究は、fMRIによって取得されたDVD動画視聴時の脳活動データを用いて時間認識時の脳活動状態を推定する。動画中の発話を文章に書き起こし、文の意味を汎用言語モデルBERT [1]を用いてベクトルとして表現し、それを言語の特徴量とみなす。言語の特徴量から脳活動データを予測する符号化モデル(2.3節参照)を構築し、言語刺激下の脳内状態を推定可能とし、その結果を可視化する。

2.1 符号化モデル

符号化モデルの構築方法として、ヒト脳への刺激となるデータから抽出した特徴量と刺激下の脳活動状態を線形回帰し、計測脳活動パターンと予測脳活動パターンが近づくように重みを学習する。

2.2 使用脳活動データ

実験に使用した脳活動データは、情報通信研究機構脳情報通信融合センターにおいて取得されたものであり、日本人被験者4人が、5本の映画やドラマのDVDを自由視の指示のもと視聴している脳活動を観測したものである。映画やドラマの内訳は4本が海外の映画またはドラマであり、残り1本は日本のアニメーションとなっている。動画視聴時の脳活動は磁気共鳴機能画像法(fMRI)を用いて脳内の血中酸素濃度依存型信号を観測したものである。脳活動データは、1秒間隔で計測され、時点毎の観測ボリュームは $96 \times 96 \times 72 (= 663, 552)$ ボクセルとなっている。

2.3 時間特徴量抽出のための深層学習モデル

DVD視聴時に発話内に現れる時間概念を識別するための深層学習モデルを構築する。これは、その識別モデルの中間層に表される特徴量(表現ベクトル)は、時間概念を識別するための表現になっていると考えることができるため、その特徴量から符号化モデルにより推定した脳活動状態によって、時間概念を捉えるための脳内の特性を解明するためである。時間識別深層学習モデルが識別する時制は、「過去」「現在」「未来」「その他」の4つとなっている。

発話内容の特徴量は、日本語用のNWJC-BERT [2]を使用し、発話を構成する1文ごとにBERTの潜在トークンである[CLS]の埋込ベクトルを利用した。モデルの概要図を図1に示す。

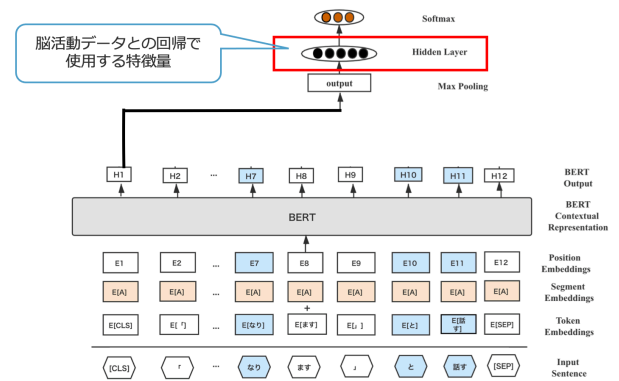


図 1: 時間識別モデル概要図

2.3.1 多段階ファインチューニング

本研究では、時間識別モデルの推定精度向上のため、類似のタスクを学習したモデルを多段に転移してモデルを洗練させていく転移学習手法である、マルチステップファインチューニングを適用した。今回は、目標タスクをDVD視聴時の発話内の時間概念の識別とし、1段階目に類似した時間関係タスクとして、本研究では、ある一つの話題について一人が600~700秒ほどスピーチしている日本語話言葉コーパス内の時間概念を識別するタスクで学習させたモデルを採用し、2段階目にDVD視聴時の時間概念識別のタスクを解いた。

2.4 時間識別特徴量のみに基づく脳内状態予測

作成した時間識別モデルの上位層の特徴量から事前学習済みBERTの埋込ベクトルが入力される低層の純粋な言語特徴量を取り除くと、純粋に時間識別のための特徴量が得られるとの考えから、それぞれの特徴量から推定された脳活動状態の差異を観測した。この際、それぞれの特徴量から個別に脳内状態を推定してしまうと、それぞれを共通の空間における比較にはならないため、双方の特徴量を合わせて回帰を行う。Banded Ridge Regression [3]を用いて、それぞれの特徴量から予測された脳内状態の差異をとれるようにした。

3 実験

2.3節で説明した時間識別深層学習モデルを用いて、その中間層から時間識別のための特徴量を抽出し、実際に観測された脳活動データとの回帰モデルを構築する。図2にその概要を示す。

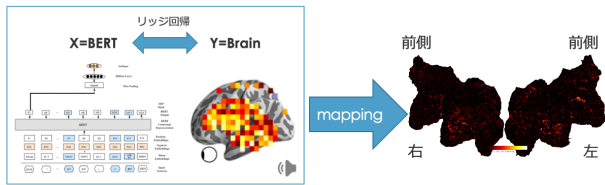


図 2: 符号化モデル作成の概要図

3.1 実験設定

モデル構築に使用したデータは、5本のDVD視聴データの内、4本の動画を訓練データとして使用し、残りの1本を評価データとして構築した。また、BERTにより表現した発話内容の特徴量からfMRIにて取得された脳活動データにより表現される脳内状態を推定するための符号化モデルとして、本研究ではRidge線形回帰を採用する。回帰の際は、被験者が動画を見てからfMRIで観測される脳活動に影響が出るまでに生じる時間のずれを考慮し、脳活動データと時間識別モデル特徴量を4秒ずらして対応づけた。また、この際、前後の刺激に対する応答が重なるため、ある時点 t の脳活動について、書起特徴の複数時点が対応するように、BERTの埋め込みベクトルを結合した行列を使用した。訓練データと評価データは観測した脳活動データの連続性を考慮して分割し、訓練データに対する5分割交差検証を行うことにより、脳活動データのボクセル毎に最適な正則化項を決定した。回帰モデルにより予測されたデータと、実際に観測された脳活動データ(正解データ)をピアソンの積率相関係数により評価を行う。またこの際、帰無仮説「推定値と評価値には相関関係が存在しない」を仮定のうえ、各ボクセルごとに推定した値に対し、 $p = 0.05$ の下で両側検定、この仮説を棄却した信用できるボクセルのみ使用している。

3.2 実験結果

図3(a)にDVDの発話を時間識別モデルの入力とした際の時間識別特徴量からの予測脳活動状態を示す。図3(b)に多段階ファインチューニングを行い脳内状態推定精度を向上させた際の予測脳活動状態を示す。さらに、より詳細に時間識別領域についてみるため、時間に関する特徴量から純粋な言語特徴量を引いた際の予測脳活動状態を図3(d)に示す。

3.3 考察

図3(a)のDVD発話データのみを用いて時間識別モデル構築を行なった際には、予測脳活動で特徴的な反応を見ることができなかったが、図3(b)においては予測された領域の局在性が見えるようになり、後頭葉の部分に強い反応が見られた。この部分は主に言語を司る部分だと言われており、反応としては妥当であると言えるであろう。この結果から多段階ファインチューニングを行うことで、モデルの精度がより良くなり、予測脳活動の精度がよくなったことがわかった。また、時間識別モデルから得た特徴量から図3(c)に示したBERT低層の言語特徴量を引いた図3(d)においては、全被験者を通して前頭前野の反応がよく残る傾向となった。このことからこの部分で何かしらの時間処理が行われている可能性が考えられる。



図 3: 実験結果

4 おわりに

本研究は、ヒト脳内において自然言語文中に表される時制がどの部位において処理がなされているかを解明することを目的とし、DVD視聴時の時制を伴う発話文を刺激として用い、予測した脳活動状態から時間認識に特化した脳内部位を調査した。脳内推定には、予測精度の高い符号化モデルを作成する必要があるが本研究では、深層学習モデル学習時の多段階ファインチューニングにより予測精度が向上することが確認できた。これにより、時間識別深層学習モデルを通じて得られた時間識別のための特徴量から脳内状態を推定することで、時間認識をしている部位の特定に努めた。実験を通じて判定された脳内時間処理の部位については、映画内容や被験者により多少の違いがでたものの、前頭前野の位置で反応がよく見られる傾向にあった。しかし、未だその部位が特定できたとは言い難い。今後は、脳内における時間処理の部位を詳細に解明するため、符号化モデルの改良や、実験データをより増やすことなどし、引き続き調査を進めていくつもりである。

参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc. of NAACL2019*, pp. 4171–4186, June 2019.
- [2] 浅原正幸, 西内沙恵, 加藤祥. Nwjc-bert: 多義語に対するヒトと文脈化単語埋め込みの類似性判断の対照分析. 言語処理学会第26回年次大会発表論文集, pp. 961–964, 3 2020.
- [3] Tom Dupré la Tour, Michael Eickenberg, Anwar O Nunez-Elizalde, and Jack L Gallant. Feature-space selection with banded ridge regression. *bioRxiv*, 2022.