ヒト脳における時間認識時の脳内状態の推定

[1] 須藤百香 [4] 小出（間島）直子 [3] 浅原 正幸 [2,4] 山口 裕人
[4] 久保 理恵子 [2,4] 西本 伸志 [1] 小林 一郎
[1] お茶の水女子大学 [2] 大阪大学 [3] 国立国語研究所 [4] 情報通信研究機構
[g1720523@is.ocha.ac.jp],[naoko-ko@nict.go.jp]
[masayu-a@ninjal.ac.jp],[hyamaguchi@nict.go.jp],[rkubo@fbs.osaka-u.ac.jp]
[nishimoto.shinji.fbs@osaka-u.ac.jp],[koba@is.ocha.ac.jp]

概要

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．このような背景を踏まえて，本研究では被験者が DVD を視聴している際の脳活動状態を計測し，動画の発話によって与えられる言語刺激からヒト脳内における時間認識時の脳内状態を調査することを目的とする．

1 はじめに

”時間”という概念が，ヒト脳内にどのように存在するのか，またはどの部位で処理されているのかという疑問に対する明確な解は未だに存在しない．時間は物理現象として捉えられ，我々が感覚的に感じるものであったり，文章中に記述されている事象の生起の順序関係から論理的に考えられたりと，様々な捉え方がある．近年，深層学習や汎用言語モデルといった革新的な技術の発展により，脳神経科学において脳内情報処理を解明するためにそれらの技術が頻繁に利用されるようになっている．とくに，Yamins ら [1] によって視覚情報を処理する深層学習モデルの層と脳内の視覚情報処理の階層的処理の層の間に相同性があることが示されて以来，深層学習モデルを脳の機能を理解するための作業モデルとして利用するようになっている．とくに，脳に与えられる言語刺激を表現する際に，言語の特徴量を word2vec [2] や BERT [3]，GPT-2 [4] などの汎用言語モデルを用いて表現し利用されるようになっている．Schrimpf ら [5] は脳活動データから単語の推定を行う課題において，43 の汎用言語モデルを用いてその精度を検証し，該当タスクにおいてどの言語モデルが脳活動を表現するのに適しているのかなどを調査している．このよう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になってきている．

このような背景を踏まえて，本研究では，自然言語文章中に表されている時間的概念がヒト脳内において認識されている状態はどのようなものであるかを調査することを目的とする．自然言語文章から時間を識別する深層学習モデルを構築し，ヒト脳に与えられる言語刺激から時間識別のための特徴量を抽出し，それを持って脳内状態を予測する．これにより，時間識別の脳内状態の調査を試みる．

2 時間認識時の脳活動状態推定

2.1 研究概要

図 1 に研究の概要を示す．

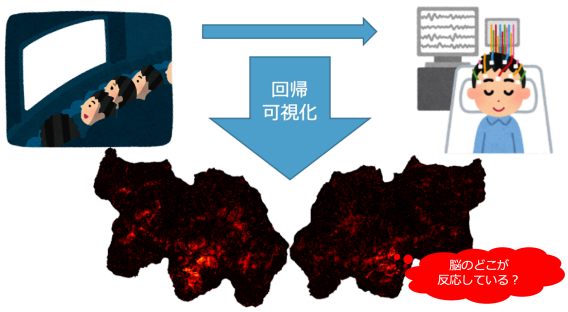


図 1: 言語刺激による脳活動状態推定の概要

DVD 動画視聴時の脳活動データを fMRI を用いて計測する．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 [3] 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（2.2 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．

2.2 符号化モデル

本研究における符号化モデル（Encoding model）の構築手法は，Naselaris ら [6] によるものを採用した．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．

2.3 被験者実験

被験者は日本人 4 名で，5 本の映画やドラマの DVD を自由視の指示のもと視聴した．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である．海外映画およびドラマとも日本語に吹き替えされたものであり，被験者は日本語によって動画を理解している．実験は，情報通信研究機構脳情報通信融合センターにおいて実施され，上記 DVD データに関して，約 9 時間の動画視聴時の脳活動を磁気共鳴機能画像法（functional magnetic resonance imaging:fMRI）を用いて脳内の血中酸素濃度依存型（BOLD:blood oxygenation level dependent）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 $96 \times 96 \times 72 (= 663,552)$ ボクセルである．

2.4 時間識別深層学習モデル

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．使用したデータの具体例を表 1 に示す．

表 1: 書き起こし文と時制の対応

時制	発話内容
⋮	⋮
現在	すみません。もうちょっとさがってください。ご協力お願いします。
その他	警部。
未来	リズボン捜査官。君たちの出番はないだろう。遺体を見つけた近所の子がホシだ。
過去	自白したんですか。
過去	ああ、ありゃ変人だよ。話を聞いたが。マーシーをころしたんですか
⋮	⋮

発話内容の特徴量は，日本語用の NWJC-BERT [7] 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 2 に示す．

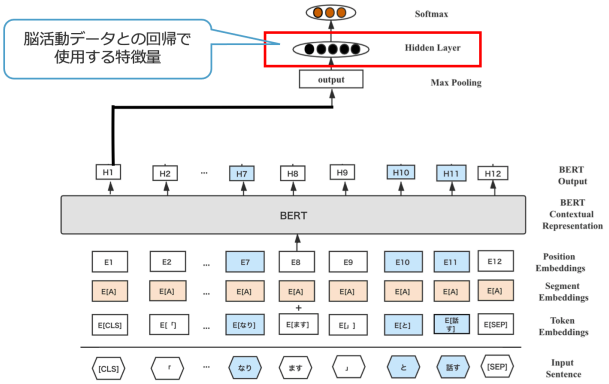


図 2: 時間識別モデル概要図

3 実験

3.1 符号化モデル作成

2.4 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 3 にその概要を示す．

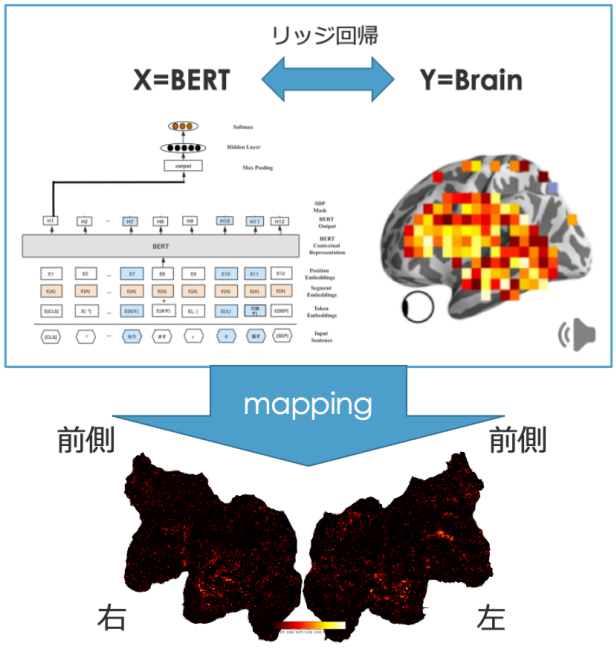


図 3: 符号化モデル作成の概要図

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，Ridge 線形回帰を採用する．また，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．

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，多段階ファインチューニング（Multi-step fine-tuning）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，日本語話し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 段目に DVD 視聴時の時間概念識別のタスクを解いた．概要を図 4 に示す．

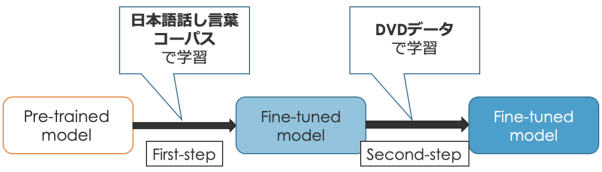


図 4: 多段階ファインチューニング

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最終層に近い上位層の特徴量から事前学習済み BERT の埋め込み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 [8] を用い，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．図 5 にその概要を示す．

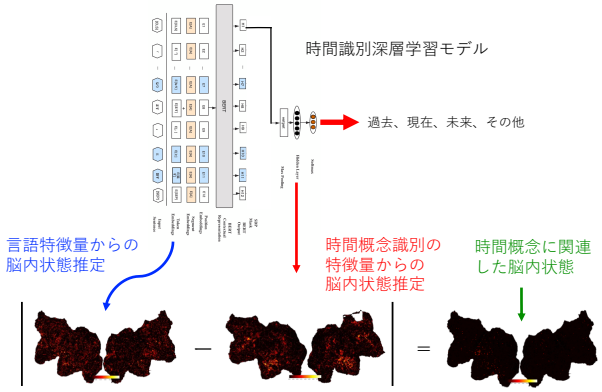


図 5: 時間識別特徴量のみに基づく脳内状態推定

3.2 実験設定

使用データ

2.3 節での被験者実験によって取得された DVD 視聴時の脳活動データとその発話の書き起こしデータ，および，日本語話言葉コーパスを傾聴した際の脳活動データとその発話の書き起こしデータを用いる．日本語話し言葉コーパスは，国立国語研究所・情報通信研究機構（旧通信総合研究所）・東京工業大学が共同開発した，日本語の自発音声を大量にあつめて多くの研究用情報を付加した話し言葉研究用のデータベースである．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ものを使用している．

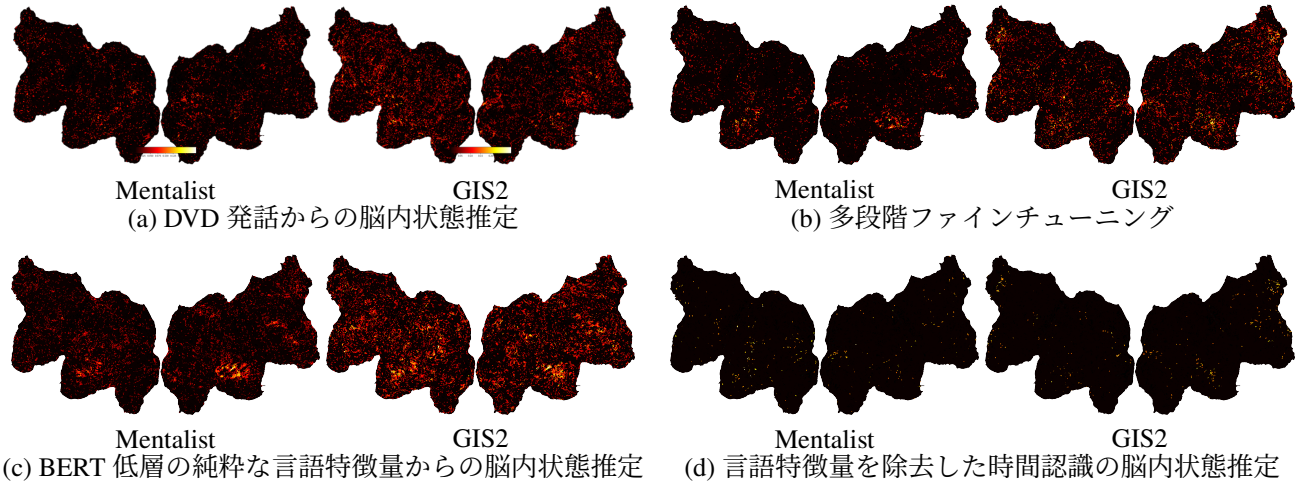
時間識別深層学習モデル構築

2.4 節で説明した時間識別深層学習モデルを構築する．モデルのパラメータを表 2 に示す．学習率や最適化アルゴリズムは先行研究 [9] を参考に採用している．

表 2: 時間識別深層学習モデルのパラメータ

学習率	埋め込みベクトル	最適化アルゴリズム
2e-5	NWJC-BERT	AdamW
エポック	バッチサイズ	隠れ層
20	16	768

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．同じ動画を訓練と評価データに分けないようにすることにより，構築するモデルが過学習を起こさないように配慮した．



(a) DVD 発話からの脳内状態推定　(b) 多段階ファインチューニング

(c) BERT 低層の純粋な言語特徴量からの脳内状態推定　(d) 言語特徴量を除去した時間認識の脳内状態推定

（上位 500 ボクセル）

図 6: 実験結果

3.3 実験結果

図 6(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 6(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 6(d) に示す．

3.4 考察

図 6(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 6(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 6(c) に示した BERT 低層の言語特徴量を引いた図 6(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．

4 まとめ

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．脳内状態を推定するためには，予測精度の高い符号化モデルを作成する必要があり，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．

謝辞

本研究は科学研究費補助金（18H05521）の支援を受けて実施されたものである．ここに謝意を表す．

参考文献

[1] D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. **Proceedings of the National Academy of Sciences**, pp. 8619 – 8624, 10/2014 2014.

[2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, **Advances in Neural Information Processing Systems**, Vol. 26, pp. 3111–3119, 2013.

[3] 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.

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems**, Vol. 33, pp. 1877–1901. Curran Associates, Inc., 2020.

[5] Martin Schrimpf, Idan Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini, Nancy Kanwisher, Joshua Tenenbaum, and Evelina Fedorenko. The neural architecture of language: Integrative reverse-engineering converges on a model for predictive processing. **bioRxiv**, 2020.

[6] Thomas Naselaris, Kendrick N. Kay, Shinji Nishimoto, and Jack L. Gallant. Encoding and decoding in fmri. **NeuroImage**, Vol. 56, No. 2, pp. 400–410, May 2011.

[7] 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.

[8] Tom Dupré la Tour, Michael Eickenberg, Anwar O Nunez-Elizalde, and Jack L Gallant. Feature-space selection with banded ridge regression. **bioRxiv**, 2022.

[9] 耿晨婧, 程飛, KANASHIRO Pereira Lis, 浅原正幸, 小林一郎. 依存関係と文脈表現を用いた日本語時間関係識別. 人工知能学会全国大会論文集, Vol. JSAI2020, pp. 4Rin115–4Rin115, 2020.