

異被験者間の脳活動データにおける相互変換への取り組み

張 嘉瑩 (指導教員：小林 一郎)

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 [1, 2]。また、近年の深層学習の成果を取り入れて脳活動データを解読する研究も報告されている。松尾ら [3] は画像刺激を受けた際の脳活動データを入力としてその人が想起している意味情報、すなわち画像によって想起された事象を深層学習を用いて自然言語文で説明する手法を提案している。そのような深層学習を用いて脳活動データを取り扱う際、学習のために大量のデータが必要である。一方、fMRI を用いた脳活動データの収集コストが高いことや、脳のサイズに個人差があるためにたとえ同じ課題の実験であったとしても同じ被験者の脳活動データしか使用できないことにより、機械学習において使用できるデータ数が不足するといった問題が生じる。これらの背景を踏まえて、本研究では、異なる被験者の脳活動データ同士の対応関係を学習し、一人の被験者の脳活動データを擬似的に増やし、それを深層学習に使用できるようにすることを目指す。

2 提案手法

異なる被験者同士の脳活動データを統一する際に、異なる被験者から取得される脳活動データそれぞれをマルチモーダルデータとみなし、マルチモーダル情報を相互に変換する手法を適用して、ある被験者の脳活動データを他の被験者の脳活動データから擬似的に作り出すことを考える。そして、擬似的に増やしたデータを使用できるようにすることを目指す。

2.1 BiDNN

本研究では、Vukotic らによって提案された Bidirectional Deep Neural Network (BiDNN) を用いる。BiDNN では、一つ目のモダリティを入力、二つ目のモダリティを出力とするネットワークと、入出力を逆にした対称的なネットワークの二つを別々に学習させる。損失関数は平均二乗誤差とする。その際、それぞれのネットワークの中間層を重みを転置して共有させることにより、複数モーダルの対応関係を取ることができる。

2.2 提案手法の概要

図 1 に提案手法の概要を示す。具体的には、使用する脳活動データを、動画像タスク 1 のみを視聴した被験者 A と動画像タスク 1 と 2 を視聴した被験者 B のものと仮定して、以下の手続きを行う。

step 1. 学習

BiDNN を用いて、動画像タスク 1 を視聴したときの A と B の脳活動データをそれぞれ入出力とし、対応関係を学習する。

step 2. 変換

step 1. におけるモデルを使用し、学習に用いたタスク 1 とは別の動画像タスク 2 を視聴した B の脳活動データを、A の脳活動データに変換する。これにより、擬似的に動画像タスク 2 を視聴した A の脳活動データが

得られる。

step 3. 評価

評価においては、松尾らが提案した手法 [3] を利用する。動画像タスクを視聴しているときの脳活動データを入力とし、その際、被験者が想起している意味情報の説明文を出力する。動画像タスク 1 のみを視聴した被験者の脳活動データを入力としたときと比べて、動画像タスク 1 と 2 の両方を視聴した被験者の脳活動データを入力としたときに出力された説明文の精度が上がっているのか確認する。

3 実験

3.1 学習

まず、Vukotic ら [4] によって提案された BiDNN を構築し、異なる被験者の脳活動データ同士の対応関係を学習した。

使用するデータは、動画像タスク 1 を被験者に視聴させたときの血中酸素飽和度信号を、fMRI を用いて記録した脳活動データである。被験者 A の脳活動データは、脳活動の観測領域のうち皮質に相当する 62,552 次元のデータを、被験者 B の脳活動データは、脳活動の観測領域のうち皮質に相当する 70,933 次元のデータを利用している。データ数は、train が 6,900 サンプル、test が 300 サンプルである。

深層学習のフレームワークである Chainer¹ で実装した。まず、現有する計算機の性能では入出力に約 60,000 次元のデータが使えなかったため、AutoEncoder を用いて二人の被験者の脳活動データをそれぞれ 7,500 次元に圧縮した。学習アルゴリズムは確率的勾配降下法、損失関数は平均二乗誤差を用いている。次に、BiDNN を用いて二人の被験者同士の脳活動データを学習した。学習のユニット数は 7,500 - 5,500 - 2,000 - 5,500 - 7,500 とし、学習アルゴリズムは確率的勾配降下法、損失関数は平均二乗誤差を用いている。

3.2 変換

次に、実験 1 で学習したモデルを使用して、データの変換をした。

実験 1 とは異なる動画像タスク 2 を被験者に視聴させたときの血中酸素飽和度信号を記録した脳活動データを使用した。こちらは被験者 B のみとなり、70,933 次元のデータである。データ数は 9,000 サンプルである。

まず、入出力に約 60,000 次元のデータが使えないので、実験 1 と同様に AutoEncoder を用いて 7,500 次元に圧縮した。そして、実験 1 で学習したモデルを使用して、データの変換をした。これにより、被験者 A が擬似的に動画像タスク 2 を見たとする脳活動データが得られる。このデータの次元数は 7,500 で、データ数は 9,000 サンプルである。

¹<https://chainer.org/>

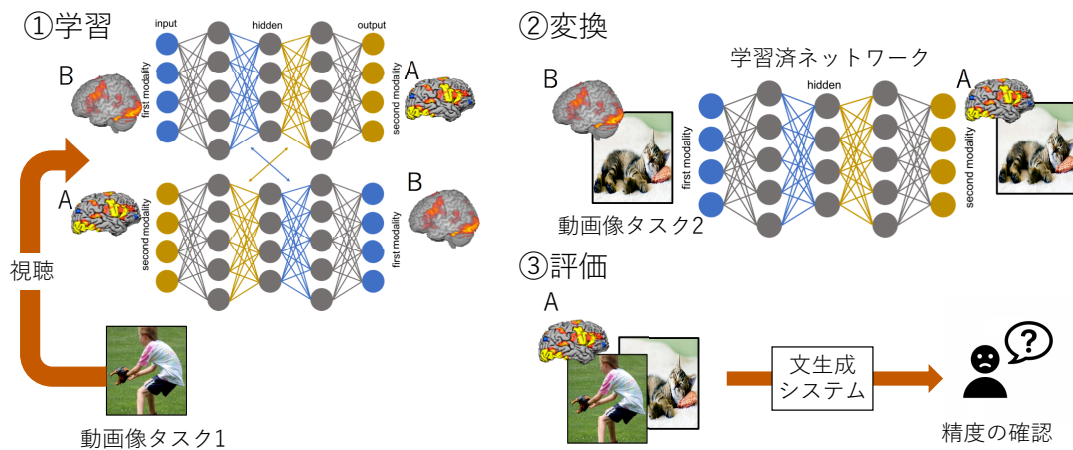


図 1: 本研究の概要

3.3 評価

最後に、松尾らが提案した手法 [3] を利用して評価をする。まず、動画像視聴時に見ていたとされる動画像から切り出した静止画に対して caption 生成システムを用いて生成した説明文を正解文とする。次に、動画像視聴時の脳活動データを入力とし、その時想起している言語意味情報の説明文と先ほどの正解文を BLEU スコアで評価する。

正解文の生成に用いる静止画のデータ数は、動画像タスク 1 から切り出したものが 7,200 サンプル、動画像タスク 2 から切り出したものが 9,000 サンプルである。また、脳活動データは被験者 A に動画像タスク 1 を視聴させたときに観測した 7,200 サンプルと実験 2 より擬似的に得られた 9,000 サンプルと、被験者 B に動画像タスク 1 と動画像タスク 2 を視聴させたときにそれぞれ観測した 7,200 サンプルと 9,000 サンプルである。

評価実験を 2 つ行う。まず、被験者 A のデータのうち、動画像タスク 1 を視聴させたときの 7,200 サンプルを使用したときと擬似的に動画像タスク 2 を視聴させたときを加えた 16,200 サンプルを使用したときの精度を計測する実験である。これにより、擬似的に 9,000 サンプル増やした際に精度にどのような影響を与えるのかを考察する。次に、被験者 B データのうち、動画像タスク 1 を視聴させたときの 7,200 サンプルを使用したときと動画像タスク 2 を視聴させたときを加えた 16,200 サンプルを使用したときの精度を計測する実験である。これにより、実際に視聴している 9,000 サンプルを増やした際に精度にどのような影響を与えるのかを考察できる。

上記二つの評価実験における BLEU スコアを表 1 に示す。BLEU スコアは、train データと test データの双方について求めた。

表 1: 評価実験における BLEU スコア

被験者	タスク	BLEU スコア (train / test)
A	1	0.5026 / 0.5027
	1, 2(擬似)	0.5238 / 0.5099
B	1	0.5028 / 0.5085
	1, 2	0.5191 / 0.5086

3.4 考察

被験者 A における評価実験の結果を見ると、train と test のどちらも擬似的なデータを増やした際にはわずかに BLEU スコアが上がっている。また、被験者 B における評価実験の結果を見ても、train と test のどちらもデータを増やした際にはわずかに BLEU スコ

アが上がっている。当初予想していたほどスコアは上がらなかったが、擬似的ではなく実際にタスク 2 を視聴した際も BLEU スコアが大きくは上がらないことから、提案手法が機能していないとは決して言えない。その原因の一つに、使用したデータが視聴している動画に問題があることが考えられる。実際に、使用した動画の中には図 2 のようなものと比べて図 3 のように、直感的に説明文をつけることが難しいものが含まれている。

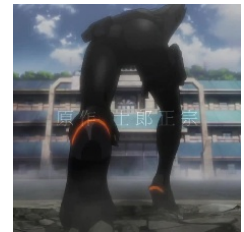


図 2: 説明しやすい画像例 図 3: 説明しづらい画像例

4 おわりに

本研究では、異なる被験者同士の脳活動データを相互変換するために BiDNN を用いた手法を提案した。脳活動データから被験者が想起している言語意味情報を説明文として出力することにより、提案手法の精度を確認し、一応の精度向上が見られたことから、概ね提案手法は有効だと考えられる。今後の課題として、直感的に説明文をつけやすい動画のみを使用した際に精度がどのように変わるかを確認するなどを行い、本研究の提案手法を再度検討したい。

参考文献

- [1] Cukur, S. Nishimoto, A. G. Hut, and J. L. Gallant, "Attention during natural vision warps semantic representation across the human brain", *Nature Neuroscience* 16, 2013.
- [2] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant., "Reconstructing visual experiences from brain activity evoked by natural movies", *Current Biology*, 21(19), 1641-1646, 2011.
- [3] 松尾映里, 小林一郎, 西本伸志, 西田知史, 麻生英樹, "画像説明文生成手法を援用した画像刺激時の脳活動の説明文生成", *言語処理学会*, P6-2, 2017.
- [4] V. Vukotic, C. Raymond, G. Gravier, "Bidirectional Joint Representation Learning with Symmetrical Deep Neural Networks for Multimodal and Crossmodal Applications", *the 2016 ACM International Conference on Multimedia Retrieval (ICMR)*, pages 343-346, ACM, New York, 2016.