

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[†] 小林一郎[†] 松尾映里[†] 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

[†]{g1420526,koba,g1220535}@is.ocha.ac.jp, [‡]{nishimoto,s-nishida}@nict.go.jp,
[¶]h.asoh@aist.go.jp

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, 4]。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 [5]。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、functional Magnetic Resonance Imaging (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、機械学習のために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つの形に統一することで、データを擬似的に増やし、それを実際に深層学習に使用できるようにすることを目指す。

2 BiDNN

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。これに対して、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。複数モーダルを学習できるものに、Ngiam らによる Multimodal Autoencoders [3] もあるが、BiDNNの方が精度が良

いと報告されている [6] ことから BiDNN を用いることにする。

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、二つ目のモダリティを入力、一つ目のモダリティを出力とする先ほどとは別のネットワークの二つを図 1 のように別々に学習させる。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 1 のように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。

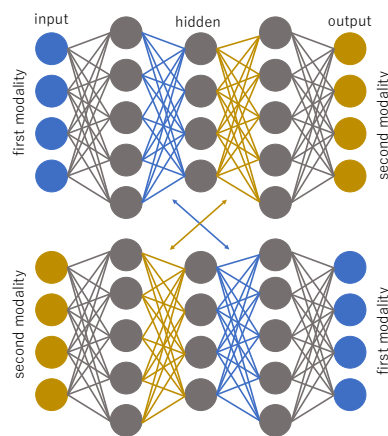


図 1: BiDNN のモデル

3 提案手法

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て統一することで、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。図 2 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。

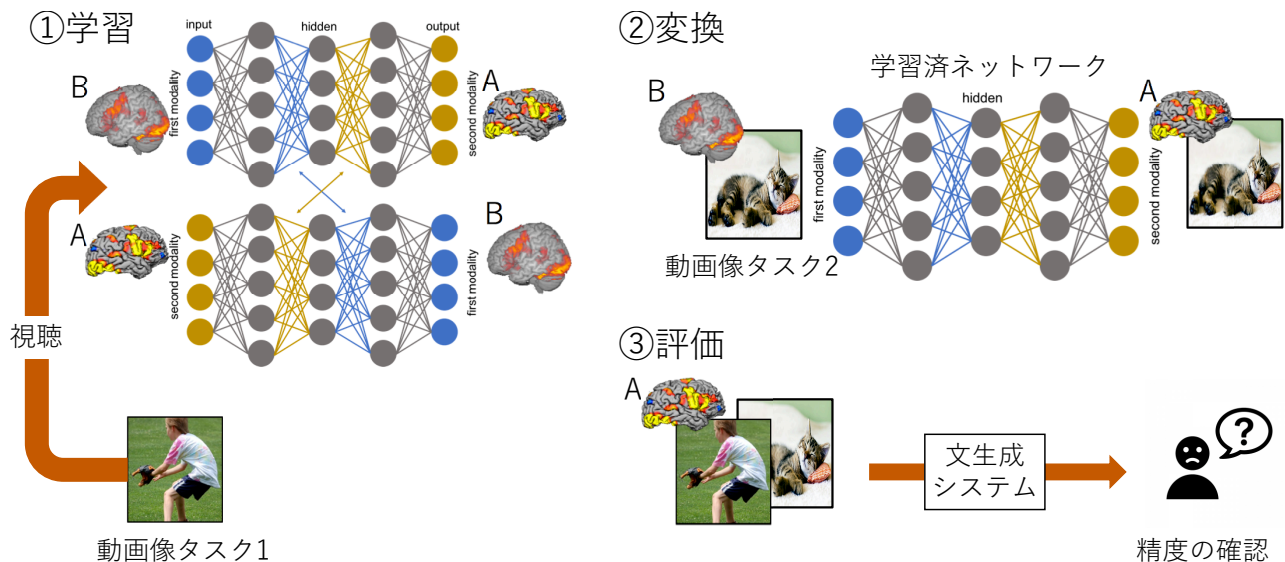


図 2: 本研究の概要

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 [5] 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。

4 実験

4.1 学習

本研究では、まず Vukotic ら [6] 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。

4.1.1 データ

使用するデータは、動画像タスク 1 を被験者に見せたときの血中酸素飽和度信号 (BOLD 信号) を。fMRI を用いて、2 秒で 1 サンプル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 $96 \times 96 \times 72$ ボクセルのうち、皮質に相当する 65,665 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 $96 \times 96 \times 72$ ボクセルのうち、皮質に相当する 68,942 次元のデータを利用している。データ数は、train が 4,200 サンプル、test が 300 サンプルとしている。

4.1.2 実験設定

先行研究 [6] では、深層学習のフレームワーク Theano を利用しているが、本研究では同じ深層学習のフレームワークである Chainer で実装した。

まず、メモリの問題で入出力に約 60,000 次元のデータが使えなかったため、AutoEncoder を用いて次元削減をした。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。

4.2 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。

4.2.1 データ

実験1とは違う動画像タスク2を被験者に見せたときの血中酸素飽和度信号 (BOLD 信号) を2秒で1サンプル記録した脳活動データを使用した。こちらは被験者Bのみで、68,942次元のデータである。データ数は2,100サンプルである。

4.2.2 実験設定

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2,100サンプルである。

4.3 評価

最後に、松尾らが提案した手法 [5] を利用して評価をする。図3に概要図を示す。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。

4.3.1 データ

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4,500サンプル、動画像タスク2から切り出したものが2,100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を見せたときに観測した4,500サンプルと実験2より擬似的に得られた2,100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を見せたときにそれぞれ観測した4,500サンプルと2,100サンプルである。

4.3.2 実験設定

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を見せたときの4,500サンプルを使用したときと動画像タスク1を見せたときのデータと擬似的に動画像タスク2を見せたときの合わせて6,600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2,100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が見ることができる。次に、被験者Bデータのうち、動画像タスク1を見せたときの4,500サンプルを使用したときと動画像タスク1と2を見せたときの合わせて6,600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2,100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することができる。

4.3.3 実験結果

評価実験2つにおけるBLEUスコアを表1に示した。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	0.5067 / 0.5476
	1, 2(擬似)	0.5055 / 0.5438
B	1	0.5290 / 0.5531
	1, 2	0.5180 / 0.5448

4.4 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スコアが下がっている。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スコアが下がっている。データを増やしたにも関わらず、当初予想していた結果とは逆の結果となってしまったが、擬似的ではなく実際にタスク2を視聴した際のBLEUスコアの減少と、擬似的にタスク2を視聴した際のBLEUスコアの減少の具合が似ていることから、今回の提案手法が明らかに間違っているとは言い難い。また、実際にタスク2を視聴したデータを増や

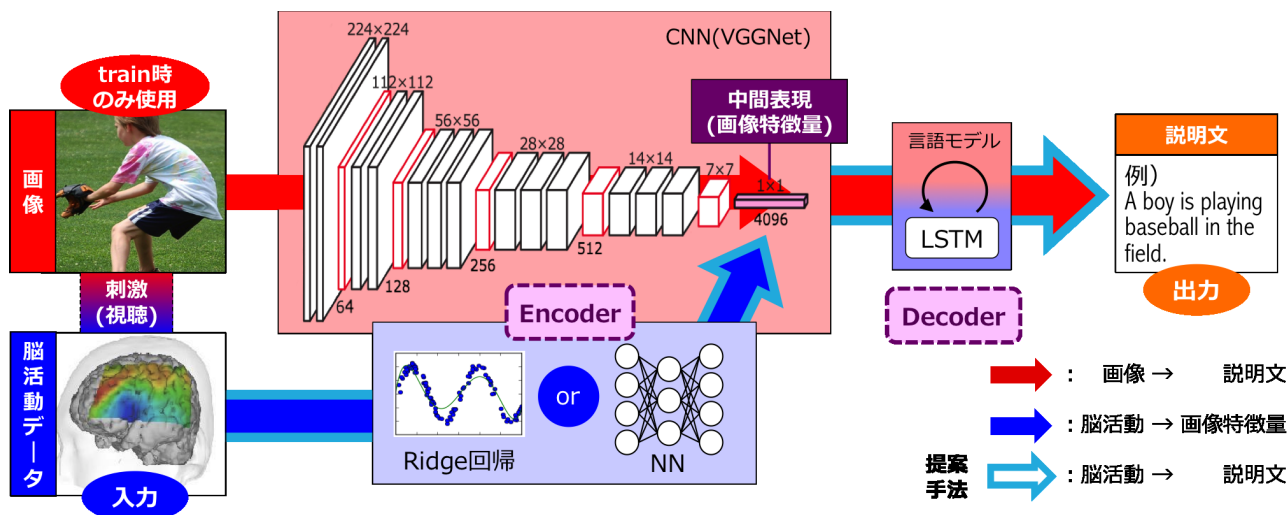


図 3: 脳活動データからの文生成の概要 [5]

した際に BLEU スコアが下がることは、図 4 のようなタスク 1 と比べて図 5 のように、視聴した動画に直感的にも説明文をつけることが難しいものが多かったこともその一因として考えられる。



図 4: タスク 1 における例



図 5: タスク 2 における例

5 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たが、今回の提案手法が有効ともそうでないとも言い難い結果となった。

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確認したり、動画画像タスク 1 のみを 2 つに分けて 2 つのタスクとみなし、今回と同じ実験をした際に精度がどのように変わるか確認したい。また、新たに動画画像タスクにおける脳活動データを追加して、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] A. G. Huth, S. Nishimoto, A. T. Vu, and J. L. Gallant, “A continuous semantic space describes the representation of thousands of object and action categories across the human brain”, *Neuron*, 76(6), 1210-1224, 2012.
- [3] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, “Multimodal deep learning”, In *Intl. Conf. on Machine Learning*, 2011.
- [4] 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.
- [5] 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.
- [6] 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.