# 画像刺激時の脳活動データを用いた脳内情報解読への取り組み

張 嘉瑩　　小林 一郎
お茶の水女子大学
{g1420526, koba}@is.ocha.ac.jp

## 1 はじめに

近年，脳神経科学分野において，脳神経活動によって作り出される意味的な情報（「意味表象」と呼ぶ）を定量的に理解する研究が盛んに行われている．とくに，近年の深層学習の発展に基づき，それらのモデルを利用した脳内情報解読の研究が進められている [2, 3, 4]．本研究では，画像刺激によってヒト脳内において想起されている意味表象を functional Magnetic Resonance Imaging (fMRI) を用いて観測し，深層学習モデルを援用することによりその観測データを表現する自然言語文を生成する手法を提案する．具体的には，一般に，fMRI を用いた脳活動データの収集コストが大きく，また脳のサイズに個人差があるために大規模なデータ収集は困難であることから，事前に学習された image captioning を行う深層学習モデルを援用することで少量データを効率的に活用し，脳活動データから自然言語文を生成する手法を開発する．

## 2 関連研究

脳活動データを用いて人が知覚している意味情報を解析する手法は複数の先行研究において報告されている [2, 3, 4] ．Çukur ら [2]，Huth ら [3] らは，動画像中の物体に注目し，ラベル分類に基づき単語レベルの意味表象推定を対象とした分析を行った．松尾ら [4] は，動画像刺激時における脳神経活動から，深層学習の手法を用いてより説明力の高い自然言語文によって意味情報を表現した．その際，image captioning 手法 [6] を援用することによって，観測コストが高く大規模なデータ収集が困難な fMRI データの効果的活用を行なった．本研究では，松尾ら [4] の提案したモデルに対し，先行研究とは異なる画像刺激時の脳活動データ（BOLD5000[5]）を適用することで，視覚刺激時における更なる脳内情報解読手法の開発を目指す．

## 3 説明文生成

松尾ら [4] のモデルを構築し，画像刺激の脳活動データにデータセットを変更して説明文生成を行う．このモデルでは脳活動データとその時視聴していた（動）画像の対応関係を学習するが，今回使用するデータでは，被験者に十分時間をあけて画像を視聴させていることから画像特徴量と脳活動データとの対応付けがより明確になっていると期待される．概要は図 1 のピンクと水色の部分である．

### 3.1 実験 1 : 画像→特徴量→説明文モデル

#### 3.1.1 実験設定

学習のためのデータセットとして，静止画とその説明文のペアからなる Microsoft COCO[1]を使用した．訓練データは 414,113 サンプル，評価データは 202,654 サンプルある．学習に関する詳細設定は表 1 の左列に示す．また，評価データとして，後に説明する脳活動データで刺激画像として与えられた ImageNet[2] と SUN[3]も使用する．

#### 3.1.2 結果と考察

test データの画像入力による文生成の結果例を表 2 に示す．学習データに用いた COCO では妥当な文が生成できたと言えるが，ImageNet と SUN は学習データに含まれていないため，文の精度はその画像が COCO に近いかどうかに依存する．ImageNet の例で示したような COCO に含まれている画像と近い場合は文の精度が良く，SUN の例で示したような遠いものでは文の精度が悪いことが分かる．

[1] http://mscoco.org/
[2] http://www.image-net.org
[3] https://vision.princeton.edu/projects/2010/SUN/

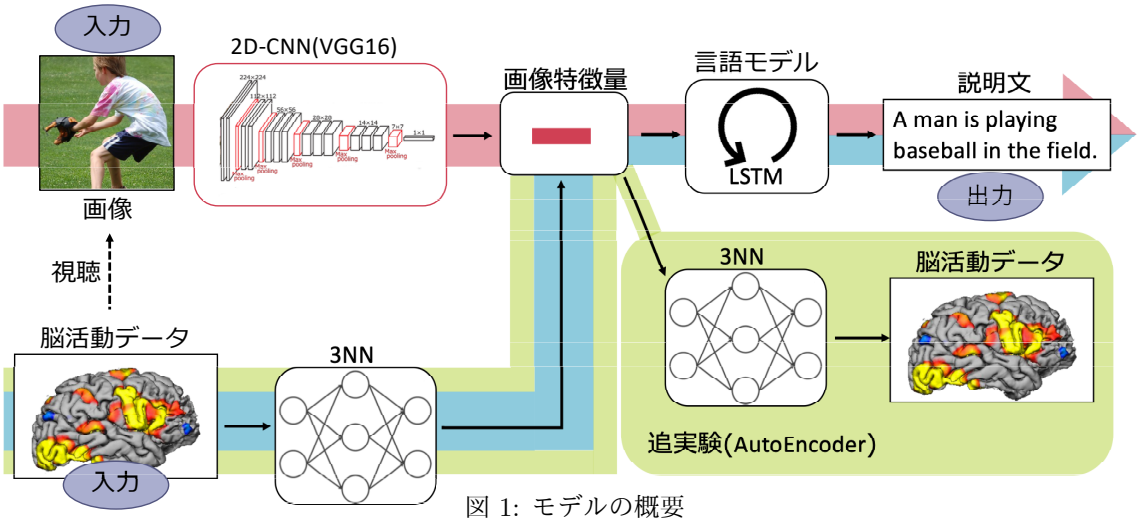


図 1: モデルの概要

表 1: 実験詳細学習設定

| | モデル 1：画像→特徴量→説明文 | モデル 2：脳活動データ→特徴量 |
|---|---|---|
| データセット | Microsoft COCO | BOLD5000 |
| データ数 (train/test) | 414,113 / 202,654 | 4,254 / 500 / 500 |
| アルゴリズム | Adam | SGD / Adam |
| 学習に関する | 学習率：0.001，勾配閾値：1 | 学習率：0.01 / 0.001，勾配閾値：1 |
| ハイパーパラメータ | L2 正則化項：0.005，epoch：100 | L2 正則化項: 0.003，epoch：100 |
| 層ユニット数 | 各層 512 | 43,312 - 10,000 - 4,096 |
| 誤差関数 | 交差エントロピー | 平均二乗誤差 |

表 2: 画像から生成した説明文の例

| | | |
|---|---|---|
| COCO |  | 0: A cat sitting on top of an open laptop.<br>1: A cat sitting on top of an open suitcase.<br>2: A cat sitting on top of an open laptop computer. |
| ImageNet |  | 0: A group of people riding motorcycles down the street.<br>1: A group of people riding bikes down the street.<br>2: A group of people riding motorcycles down the road. |
| SUN |  | 0: A motorcycle parked in front of an old building.<br>1: A motorcycle parked on the side of an empty road.<br>2: A motorcycle parked on the side of an empty street. |

## 3.2 実験 2：脳活動データ→特徴量モデル

### 3.2.1 実験設定

データセットとして，画像を被験者に視聴させた時の血中酸素飽和度信号（BOLD 信号）を fMRI を用いて 2 秒ごとに記録した脳活動データである BOLD5000[4]，およびその画像を使用する．刺激として用いた画像データセットは Microsoft COCO，ImageNet，SUN である．また，脳活動データは，皮質に相当する 43,312 次元のデータを使用し，画像から抽出した特徴量との対応関係を学習する．この時，Chang ら [1] が画像刺激を受けてからその反応が脳活動に現れるピークが 6 秒後と報告しているため，6 秒ずらして対応をとる．学習に関する詳細設定は表 1 の右列に示す．

### 3.2.2 実験結果

実際の画像特徴量とこの実験で予測した画像特徴量の相関係数を求めたものを表 3 の左列に示す．

表 3: 特徴量の相関係数

| | 実験 2（train / test） | 追実験（train / test） |
|---|---|---|
| ave | 0.197 / 0.068 | 0.352 / 0.345 |
| min | -0.229 / -0.237 | -0.113 / -0.075 |
| max | 0.883 / 0.343 | 0.703 / 0.676 |

[4]https://bold5000.github.io

表 4: 脳活動データから生成した説明文の例

| | 画像 | 実験 3（Adam） | 追実験（$\lambda = 0.1$） |
|---|---|---|---|
| train |  | A cat sitting on top of an open laptop. | A dog sitting on the ground in front of an open door. | A black and white cat sitting on top of an open window. |
| |  | A man is surfing in the ocean on his surfboard. | A group of people sitting on the ground next to each other. | A man is standing on the beach with his surfboard. |
| test |  | A tennis player is getting ready to hit the ball. | A fire hydrant sitting in the middle of some grass. | A man is sitting on the ground with his skateboard. |
| |  | A person riding skis on top of snow covered ground. | A man standing in front of an old building. | A person is sitting on the ground with an umbrella. |

## 3.3 実験 3：脳活動データ→説明文モデル

### 3.3.1 実験設定

実験 1，実験 2 で学習したモデルを組み合わせたものに対し，脳活動データを入力し説明文生成を実行する．また，その時視聴していた画像から直接モデル 1 を使用した説明文生成も行う．ここでは，先の実験 1 における結果を受け，Microsoft COCO の画像とその画像を視聴している脳活動データのみを対象に行う．

### 3.3.2 結果と考察

脳活動データと画像から生成した説明文の例を表 4 の左列と中央列に示す．各画像に対して個別の文章は出るが，画像の特徴を捉えているとは言い切れない．

## 3.4 追実験：モデルの変更

### 3.4.1 実験設定

先の実験 3 を踏まえて，精度向上のためにモデル 2 における学習時にモデルの変更を行い，3 層 NN と 5 層 AutoEncoder を組み合わせたものにする．3 層 NN は先の実験 3 と同じものを使用し，そこに 5 層 AutoEncoder での学習を加味することで 3 層目の中間層がより良い特徴表現となることを期待する．概要は図 1 の緑の部分である．学習に使用する脳活動データとその時視聴している画像の特徴量組を $(B, F)$，平均二乗誤差を $MSE(\cdot,\cdot)$，モデルの 3 層目の出力を $out_F$，5 層目の出力を $out_B$ と表すと，誤差関数は

$$loss \ = \ MSE(F, out_F) + \lambda \ MSE(B, out_B) \quad (1)$$

と書ける．詳細設定は表 5 に示す．

表 5: 追実験詳細学習設定

| | 追実験：脳活動データ→特徴量 |
|---|---|
| データセット | BOLD5000 |
| データ数 (train/test) | 4,254 / 500 / 500 |
| アルゴリズム | SGD |
| 学習に関する ハイパーパラメータ | 学習率：0.01，勾配閾値：1<br>L2 正則化項: 0.003，epoch：100 |
| 層ユニット数 | 43,312 - 10,000 - 4,096 - 10,000 - 43,312 |
| 誤差関数 | 平均二乗誤差 |

### 3.4.2 結果と考察

脳活動データから生成した説明文の例を表 4 の右列に示す．また，実際の画像特徴量と推定した画像特徴量の相関係数を表 3 の右列に示す．train については概ね画像の特徴を捉えた文章が出力されたが，test については文法は正しいものも特徴はまだ捉えられていない．原因として，学習の際に loss の値は減少したものも推測した特徴量と実際の特徴量の相関がまだ低いことが考えられる．モデルの変更により，文生成と相関係数のどちらも良くなったことが分かる．

# 4 脳状態推定

先の実験 2 とは逆に，脳活動データと対応する画像から抽出した画像特徴量から脳のボクセルの活動状態を推定することで，脳領域のどの部位が活動をしており，解剖学的な観点からそれがどのような意味なのかを分析する．

## 4.1 実験設定

まず，VGGNet により脳活動データと対応する画像から画像特徴量を抽出する．次に，その画像特徴量から脳活動データを推定する．この時，画像特徴量が 4,096 次元に対し，脳活動データが 43,312 次元とか

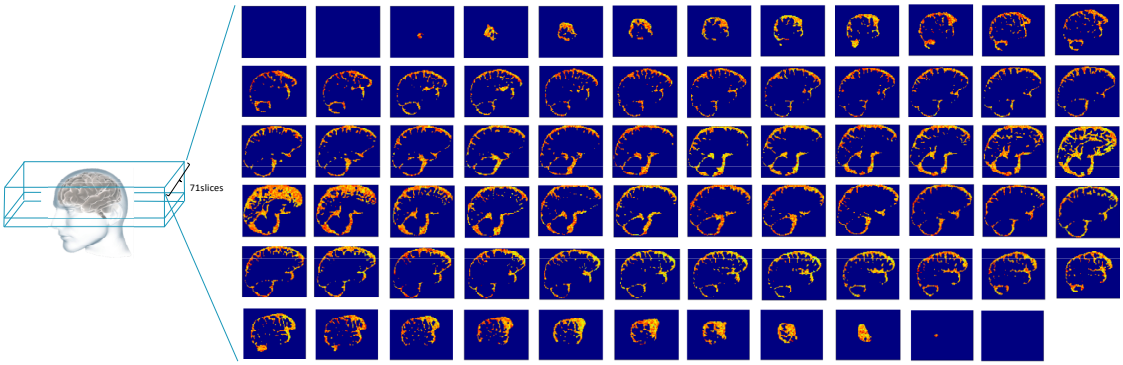


図 2: 相関係数の可視化結果

なり大きくなっており学習が難しいと判断したため，画像特徴量から脳活動データの 1 ボクセル（つまり 1 次元）を予測するモデルを 43,312 個構築して学習する．それから，脳活動データの実際のデータと予測したデータの相関係数を求める．この時，刺激時によく活動していた部位は相関係数の値も大きくなると考える．最後に，求めた相関係数を 0 から 1 に正規化した後可視化を行う．

### 4.2 結果と考察

可視化した結果を図 2 に示す．赤色の方が相関係数が高いことを表しており，各画像の右側が前頭葉，左側が後頭葉にあたる．

解剖学的に視覚刺激を受けた際の視覚経路は図 3 となっており，後頭葉付近に反応が現れると考えられるが，可視化の結果を見ると後頭葉付近も高いが，皮質全体の相関係数が高いことが分かる．はっきりとした結果は得られなかったため，他の被験者のデータを使用した再検討が必要である．また，今回の原因として，1 ボクセルずつ予測したために予測精度が上がってしまったことが考えられるので，予測範囲をもう少し広げて再検討したい．

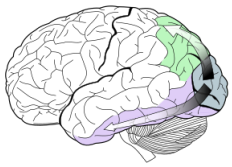

図 3: 視覚経路

## 5 おわりに

本研究では，まず，先行研究 [4] に対し，脳活動データセットとモデルの一部を変更することにより更なる分析を行った．誤差関数を変更することにより，train データについてより良く特徴を捉えられた文を生成できることを示した．次に，画像特徴量から脳状態を推定することにより，脳内における活動部位の観点から分析を行い，後頭葉の視覚野も活動しているが，全体的に値が高かったためはっきりとした結果は得られなかった．

今後の課題として，まず文生成については，脳活動データ→特徴量モデルにおいて平均二乗誤差だけではなく相関係数を取り入れることにより，値だけではなくデータの傾向も踏まえた損失関数で学習を行いたい．また，状態推定については，他の被験者の脳活動データを用いることによって，同じような結果が得られるのか再確認したい．

## 謝辞

本研究の一部は，科研費新学術領域研究（課題：18H05118）の支援を受けたものである．

## 参考文献

[1] N. Chang, J. A. Pyles, A. Marcus, A. Gupta, M. J. Tarr and E. M. Aminoff. BOLD5000, a public fMRI dataset while viewing 5000 visual images.Scientific data, 6(1), 49, 2019.

[2] T. Çukur, S. Nishimoto, A. G. Hut and J. L. Gallant. Attention during natural vision warps semantic representation across the human brain. Nature Neuroscience, 16, 763-770, 2013.

[3] 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.

[4] 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.

[5] Nadine Chang, John A. Pyles, Austin Marcus, Abhinav Gupta, Michael J. Tarr and Elissa M. Aminoff, BOLD5000, a public fMRI dataset while viewing 5000 visual images, Scientific Data volume 6, Article number:49, 2019.

[6] O. Vinyals, A. Toshev, S. Bengio and D. Erhan. Show and tell: A neural image caption generator. CVPR, 2015.