2K3-OS-33a-1in2

大脳皮質における予測符号化を模倣した動画像予測モデルと脳活動の相関に関する考察

A Study on a Correlation between a Predictive Model of Motion Pictures Imitating the Predictive Coding of the Cerebral Cortex and Brain Activity

藤山 千紘 *1 Chihiro Fujiyama　小林 一郎 *1 Ichiro Kobayashi　西本 伸志 *2 Shinji Nishimoto　西田 知史 *2 Satoshi Nishida　麻生 英樹 *3 Hideki Asoh

*1お茶の水女子大学 Ochanomizu University　*2情報通信研究機構 脳情報通信融合研究センター National Institute of Information and Communication Technology

*3産業技術総合研究所 人工知能研究センター National Institute of Advanced Industrial Science and Technology

Great strides have been made in studying deep learning architectures. In addition, developing machine learning modules imitating each organ in human brain has been actively studied as an approach to an artificial general intelligence. Here we explore a correlation between internal representations in a predictive neural network which is inspired by the concept of predictive coding from the neuroscience literature and brain activity evoked by natural movies.

1. はじめに

近年，ニューラルネットワークを用いた深層学習の技術が著しく発展し，様々なモデルが提案されている．また脳型汎用人工知能の構築を目指す取り組みの一つとして，計算機科学・神経科学双方からのアプローチにより，脳の各器官を機械学習モジュールとして開発する研究が盛んに行われている．本研究では，大脳皮質における情報処理を模倣して予測画像の生成を行う深層学習モデルを先行研究 [Lotter 16] に従って構築し，多量かつ多様な自然動画像データセットを適用して学習を行い，その有効性を確認するとともに，モデル内部における特徴表現と脳活動との相関を考察する．

2. 予測符号化

2.1 大脳皮質における予測符号化

脳は近い将来に入力されるであろう刺激を予測し，実際に入力される刺激との差分を処理することにより，効率的な情報処理を実現している．大脳皮質においては，高次領域，低次領域間に双方向の接続が存在し，高次領域で生成された予測が低次領域へと伝播，実際の入力刺激と予測の差分が低次領域から高次領域へ向かってフィードバックされることにより脳内の予測モデルが更新されるという一連の処理が行われ，より精度の高い予測を行うことが可能になるとされている．

2.2 PredNet

PredNet[Lotter 16] は，大脳皮質における予測符号化の処理を模倣し，深層学習の枠組みで構築されたモデルであり，動画像を与えられた下で将来の画像を予測するタスクを行う過程で汎用的な特徴を学習するモデルとして提案されている．PredNet では画像の特徴を掴むのに適した Convolutional Neural Network(CNN) および，系列データを扱うのに適した Long-Short Term Memory(LSTM) と CNN を結合し時空間に対して広がりを持つデータを扱えるようにしたモデルである Convolutional LSTM[Xingjian 15] を用いて動画像の特徴を掴み，予測タスクを行っている．PredNet は同じ構造を持つモジュールを深くスタックした形をとっており，一つのモジュールは，脳内の予測モデルに相当する Representation モジュール，入力処理を行う Input モジュール，予測を生成する Prediction モジュール，入力と予測との差分を生成する Error ユニットの 4 パートからなり，PredNet 全体では，上位層で生成された予測が下位層へ，下位層で生成されたエラーシグナルが上位層へ伝播するという流れで，情報の授受が行われる．

連絡先: 藤山千紘，お茶の水女子大学，〒 112-8610 東京都文京区大塚 2-1-1，g1220533@is.ocha.ac.jp

3. 実験

本研究では，初めに PredNet の構築，多量の動画像データセットを適用した学習を行い，その有効性を確認する．続いて学習されたモデルに対し，脳活動測定時に被験者に提示された刺激画像を入力として与え，その際の Representation モジュールにおける特徴表現と脳活動との対応関係を Ridge 回帰を用いて学習する．対応関係の学習後，脳活動からモデル内部の特徴表現の推定を行い，推定された特徴表現と PredNet に刺激画像を適用して得られた特徴表現との相関係数を算出する．PredNet の一部および脳活動との対応関係を図 1 に示す．

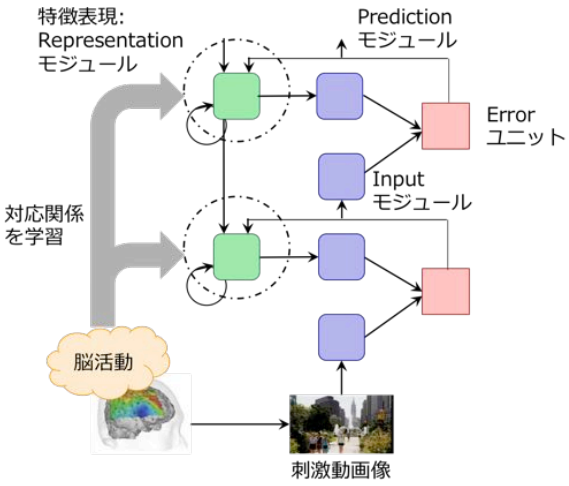

図 1: PredNet 内部の特徴表現と脳活動の対応関係

3.1 実験設定

3.1.1 PredNet の学習

PredNet の学習に関するハイパーパラメータは先行研究 [Lotter 16] の設定に基づき，層数を 4，Convolution のフィルタサイズを全て 3 × 3，フィルタ数を下位層から順に 3，48，96，192 とした．最適化アルゴリズムには Adaptive Moment Estimation(Adam) を用いた．学習は最下層 (ピクセル層) における Prediction モジュールの出力と，正解データである一時刻先の画像の各ピクセル間の平均二乗誤差の 10 フレームに亘る合計値を最小化する形で行った．学習データセットとして，参考文献 [Nishimoto 11] において用いられている自然動画像データセットを用い，動画像を 1 秒間に 10 フレームの頻度で静止画像として切り出し，160 × 120 ピクセルにダウンサンプリングした静止画像群を使用した．学習データとして 750,000 フレームを用い，5,000 フレームで 1 エポックとした上で，各エポックにおいて 500 フレームの検証データを用いて検証誤差を算出し，150 エポックに亘って学習を行った．

3.1.2 特徴表現と脳活動の対応関係

PredNet 内部の特徴表現と脳活動の対応関係の学習に際して，特徴表現は，脳活動測定時の刺激動画像を静止画像として切り出し，160 × 120 ピクセルにダウンサンプリング後，PredNet の入力とした際の PredNet 各層の Representation モジュールの出力値を最下層から順に R0，R1，R2，R3 として取り出したものを用いた．脳活動データとしては，動画像視聴時の被験者の血中酸素濃度に依存する信号 (BOLD 信号) を functional magnetic resonance imaging(fMRI) を用いて記録した脳神経活動データ 96 × 96 × 72 ボクセルのうち皮質に相当する 65,665 次元のデータを使用した．各特徴表現と脳活動データのペアを，学習データ 4,497 対，評価データ 300 対として学習を行った．Ridge 回帰学習時の正則化項の重みパラメータは 0.5 に設定した．

3.2 PredNet の学習

150 エポックに亘る検証誤差の推移を図 2 に，検証誤差最小より上位 5 モデルを表 1 に示す．また学習後評価データを用いて予測画像の生成を試みた例を図 3 に示す．

図 **2:** 学習時の検証誤差の推移

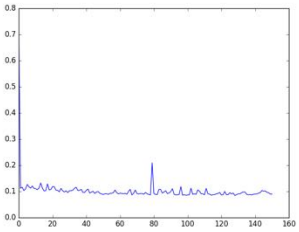

表 **1:** 検証誤差が最小の上位 5 モデル

エポック数	検証誤差
128	0.0847
99	0.0856
97	0.0857
114	0.0858
100	0.0864

学習過程について，検証誤差の推移および検証誤差が小さいモデルが 100 エポック付近にあることより学習が進んでいる様子が見受けられる．図 2 において 80 エポック付近で検証誤差の上昇が見られることに関しては，この付近の学習で用いられたデータが時間に伴う変化に乏しいデータであったことを確認しており，その影響を受けたものと推測される．生成された予測画像は図 3 上段の系列に関しては概ね一時刻先の画像を予測できていると評価できる．一方，下段の系列については予測を正確に行えていない様子が見受けられる．これは，該当評価データが学習データ中の時間に伴う平均的な変化を大きく逸するデータであったこと，予測画像生成時のエラーシグナルが

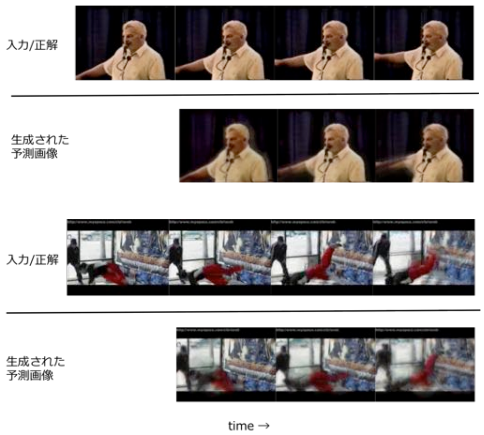

図 **3:** PredNet により生成された予測画像

十分に機能しなかったことに起因しているものと推測される．

3.3 特徴表現と脳活動の対応関係

学習した Ridge 回帰を用いて推定した各特徴表現 R0，R2，R3 と刺激画像から得られた各特徴表現の相関係数を表 2 に示す．第 1 層の特徴表現 R1 については 230,400 次元と非常に高次元であったため本稿では Ridge 回帰の学習を行わなかった．

表 **2:** 推定された特徴表現と刺激画像による特徴表現の相関係数

特徴表現	相関係数
R0	0.2490
R2	0.06695
R3	0.06984

Ridge 回帰を用いて推定された特徴表現と刺激画像入力下の特徴表現の相関係数は最下層 (R0) において 0.25 程度となり，これはノイズの多い脳活動を扱う脳神経科学分野においては，相関を認めるに値するとの知見がある．一方，より深層部に相当する R2 および R3 はほとんど相関を認められない結果となった．

4. 脳活動データを用いた画像生成

R0，R2，R3，いずれにおいても相関係数は低い結果となったが，追加実験として，学習した各 Ridge 回帰を用いて脳活動から推定した特徴表現で PredNet 内部の Representation モジュールの出力値を置換した上での画像生成を試みた．また検証のため，各特徴表現を 0 および [-1.0，1.0] のランダムな値で置換した場合の画像生成も行った．生成画像の例を図 4 に示す．

R3 に関しては，0 およびランダムな値で置換した場合は，互いにほぼ同様の画像が生成されていることが見て取れ，異なる入力を与えた場合においても結果に大きな相違をもたらさないことから，PredNet 内部において第 3 層の Representation モジュールは十分に機能していないのではないかと推測される．一方で，Ridge 回帰を用いて脳活動から推定した値で R3 を置換した場合には，0 およびランダムと比較して正解値からの生成画像に近い画像を生成できている．0 およびランダムな値で置換した場合の生成画像は類似しているが，実際には細かな差異があることを確認しており，この微小な差異を生み出すパラメータにより，推定値からの生成画像が正解値からの生成画像に近づいたものと考えられる．しかし総じて，R3 は値を置き換えても生成画像に大きな影響を及ぼすことはない，つま

図 4: 各特徴表現を置換した場合の生成画像の例．左から順に該当時刻の入力画像，Ridge 回帰を用いた脳活動からの推定値，0，[-1.0，1.0] のランダムな値で置換した生成画像，正解値からの生成画像となっている．

り PredNet 内部においては R3 からの情報は画像生成へ大きく寄与しないと考えられる．R2 に関しては，推定値で置換した場合，0 およびランダムな値で置換した場合と比較して，正解値から生成した画像に近い画像を生成できていることが分かる．R0 ではエラーシグナルからの補正を直接受け取らないため，脳活動からの推定値のみに依存して画像生成を行っており，生成画像は視認性の低いものとなった．これは，対応関係を高精度のモデルとして学習できなかったことが原因の一つであると推測される．表 3 に R0，R2，および R3 の絶対値平均と，推定値と正解値の二乗平均平方根誤差を示す．

表 3: 刺激画像による特徴表現の絶対値平均および推定値と正解値の二乗平均平方根誤差

特徴表現	絶対値平均	二乗平均平方根誤差
R0	0.2273	0.1474
R2	0.1404	0.1567
R3	0.1932	0.03864

高精度のモデルとして対応関係を学習できなかった原因としては各特徴表現の次元数が R0，R2，R3 の順に，57,600，115,200，57,600 と極めて高かったために学習が困難であったものと推測される．R0，R2，R3 はいずれも各層内では，0 およびランダムな値で置き換えた場合の生成画像に比べ，脳活動からの推定値を用いた生成画像が正解値からの生成画像に近いという傾向が見られるが，最深層部の R3 から浅い層 R0 に向かうに伴い，この傾向が顕著になっており，Representation モジュールの値の生成画像への寄与の度合いが高まっていることが分かる．

5. おわりに

本研究では，大脳皮質における予測符号化を模倣した深層予測モデルである PredNet と脳活動の相関を定量的に評価する試みとして，PredNet を構築し，その特徴表現と脳活動との相関関係を考察した．また脳活動から各特徴表現を推定し，PredNet に入力して画像生成を行い，PredNet 各層における Representation モジュールの機能を検討した．今後の課題として，脳活動・特徴表現間の対応関係としてより高精度なモデルを学習するため，PredNet 内部の特徴表現を autoencoder を用いて次元削減することや脳活動と各特徴表現の対応関係の取り方として脳活動を構造化して扱うことを検討している．

参考文献

[Lotter 16] Lotter, W., Kreiman, G., Cox, D. (2016). Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. arXiv preprint arXiv:1605.08104.

[Xingjian 15] Xingjian, S. H. I., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., Woo, W. C. (2015). ”Convolutional LSTM network: A machine learning approach for precipitation nowcasting.” In Advances in Neural Information Processing Systems (pp. 802-810).

[Nishimoto 11] Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J. L. (2011). ”Reconstructing visual experiences from brain activity evoked by natural movies.” Current Biology, 21(19), 1641-1646.