2M-09

大脳皮質における予測符号化を模倣した動画像予測に関する取り組み

藤山千紘 †　　小林一郎 ‡　　西本伸志 §　　西田知史 §　　麻生英樹 ¶

† お茶の水女子大学 理学部情報科学科　　‡ お茶の水女子大学 基幹研究院自然科学系

§ 情報通信研究機構 脳情報通信融合研究センター　　¶産業技術総合研究所 人工知能研究センター

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 [1] に従って構築し，有効性を確認するとともに，異なるデータセットを適用して学習を行い，その汎化能力を検討する．

2　予測符号化

予測符号化とは，大脳皮質で行われているとされる処理である．脳は絶えず大量の入力刺激を受け取っており，それらを効率的に処理することが求められている．先行研究 [1] では，この機能を模した深層学習モデルを構築している．本研究では，先行研究同様，予測符号化を模倣した深層学習モデルを構築し，汎化能力の考察を試みる．

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 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 Recurrent Neural Network の一種である 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 を用いて動画の特徴を掴み，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．

3　実験

本研究では，PredNet を構築し，初めに先行研究 [1] と同様の KITTI データセットを用いた学習により，モデルの有効性を確認する．続いて KITTI データセットと比してより多量であり多様性に富むデータを用いた学習を行い，PredNet の汎化能力を検討する．

3.1　実験設定

PredNet の実装に際しては，深層学習のフレームワーク Chainer を用い，学習に関するハイパーパラメータは先行研究 [1] 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 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 フレームに渡る合計値を最小化する形で行った．

3.2　KITTI データセットを用いた実験

初めに，車載カメラによって撮影された動画像を 1 秒間に 10 フレームの頻度でサンプリングした静止画像群である KITTI データセットを使用し，学習を行った．train データとしては 12,313 フレームを使用し，5,000 フレームで 1 エポックとした上で，150 エポックに渡って学習を行い，各エポックにおいて 80 フレームの validation データを用いて誤差の算出を行った．表 1 は validation 誤差が最小のモデルより上位 5 モデルを示している．

An Approach to Prediction of Motion Pictures Imitating the Predictive Coding of the Cerebral Cortex
†Chihiro FUJIYAMA(g1220533@is.ocha.ac.jp)
‡Ichiro KOBAYASHI(koba@is.ocha.ac.jp)
§Shinji NISHIMOTO(nishimoto@nict.go.jp)
§Satoshi NISHIDA(s-nishida@nict.go.jp)
¶Hideki ASOH(h.asoh@aist.go.jp)

表 1: KITTI データセットによる validation 誤差が最小の上位 5 モデル.

エポック数 (フレーム数)	validation 誤差
25 (125,000)	0.073204
16 (80,000)	0.073295
21 (105,000)	0.073355
23 (115,000)	0.073614
22 (110,000)	0.073732

3.3 自然動画像データセットを用いた実験

続いて，参考文献 [3] において用いられている動画像データを用いて学習を行った．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用いた．train データとして 750,000 フレームを用い，5,000 フレーム学習毎に 500 フレームの validation データを用いて誤差の算出を行った．その推移を図 1 に，validation 誤差最小より上位 5 モデルを表 2 に示す．また学習後，test データを用いて予測画像の生成を試みた例を図 2 に示す.

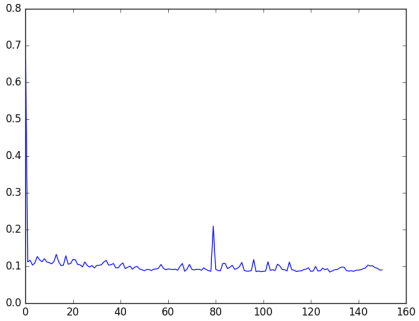

図 1: training 時の validation 誤差の変化.

表 2: 自然動画像データセットによる validation 誤差が最小の上位 5 モデル.

エポック数 (フレーム数)	validation 誤差
128 (640,000)	0.084679
99 (495,000)	0.085623
97 (485,000)	0.085654
114 (570,000)	0.085843
100 (500,000)	0.086439

3.4 考察

学習過程について，KITTI データセットを用いた実験では 25 エポック付近と早い段階において validation 誤差が最小のモデルを学習し，以降 training フレーム数を増やしても改善が見られなかった．これは train データが 12,313 フレームと小さいためと考えられる．一方，より多量で多様な自然動画像データセットを用いた実験では，validation 誤差が小さいモデルは 100 エポック付近であり，学習が進み続けている様子が見てとれる．図 1 において 80 エポック付近で validation 誤差の上昇が見られることに関しては，この付近の training 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 2 上段の系列に関して

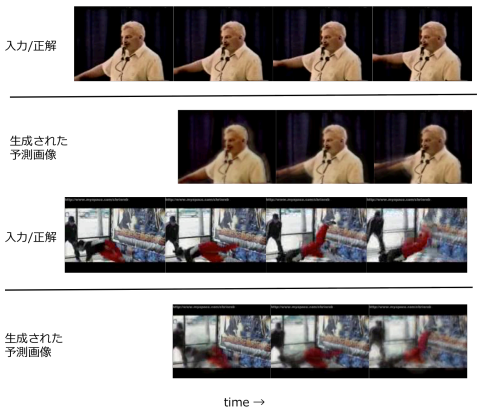

図 2: PredNet により生成された予測画像.

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，該当データが train 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．validation 誤差の改善が 0.08 付近で留まっていることに関しては，生成される画像の鮮明度の低さが原因の一つと考えられる．PredNet の汎化能力については，KITTI データセットより多様な自然動画像データセットを適用した場合においても，生成された画像からある程度の予測を行えている様子が伺え，一定の汎化能力は認められるものと考えられる.

4 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 を構築し，多様性に富む自然動画像データセットを適用して，その汎化能力を考察した．今後の課題として，PredNet と予測符号化の関連を定量的に評価するため，脳活動データと PredNet の内部表現の相関関係の考察を行うことを検討している.

謝辞

本研究では全脳アーキテクチャ研究ハッカソンにおいて開発されたソースコードを参考に PredNet を構築しました．ドワンゴ人工知能研究所の山川宏氏，ソースコードを公開された桑田純哉氏に感謝の意を表します.

参考文献

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

[2] 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).

[3] 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.