1R-07

予測符号化を模倣する深層生成学習モデル構築に向けた基礎的検討

黒田彗莉 † 西本伸志 ‡ 西田知史 ‡ 小林一郎 †

† お茶の水女子大学
‡ 情報通信研究機構 (NICT)，脳情報通信融合研究センター (CiNet)

1 はじめに

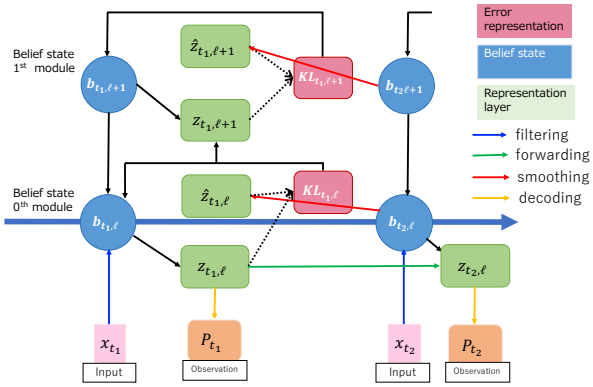
本研究では，大脳皮質で起こる予測符号化機能を模倣し，柔軟な時間間隔での映像シーンの予測を可能にするよう構築したモデル [1] の有効性を評価した．実験は，提案モデルが柔軟な時間幅で任意のシーンの予測が可能かを KITTI Benchmark データセット [2] を用いて確認する実験 (実験 1) と，構築したモデルが既存モデル（PredNet [3]）に比べて，ヒト脳内の予測符号化を説明するのに適しているかを実際に fMRI で取得した脳活動データを用いて検証する実験（実験 2）を行った．

2 提案モデル

これまでに著者らは PredNet [3] と TD-VAE [4] の機能を統合して，画像予測のための新たな深層学習モデルを構築した [1]．概要図を図 1 に示す．

PredNet は，予測符号化の概念に基づくニューラルネットワークモデルであり，大脳皮質で起こるとされているプロセスを模倣し，動画像における次の時刻の画像フレームを予測することを学習する．このモデルは次のシーンを逐次的に予測することが可能だが，柔軟な予測間隔を持つ予測を生成することはできない．一方，Temporal Differential Variational Auto-Encoder (TD-VAE) は，モデルに信念を表す状態変数を導入することで，柔軟な時間間隔での予測を可能にした深層生成モデルである．このモデルは，直接観察される状態の予測ではなく，対象がどのように振る舞うものか，という潜在的な状態の予測を行っている．

提案モデルは PredNet と TD-VAE の特性を組み合わせ，ヒト脳内での予測，つまり，ヒト脳内のような階層的な構造を持ち，自由な時間幅での予測が可能だと考える．



$$L = KL(z_{t_1,\ell}, \hat{z}_{t_1,\ell}) + KL(z_{t_1,\ell+1}, \hat{z}_{t_1,\ell+1}) + \log P(z_{t_2,\ell}|b_{t_2,\ell}) - \log P(z_{t_2,\ell}|z_{t_1,\ell}) - \log P(P_{t_2}|z_{t_2,\ell})$$

図 1: 提案モデル 概要図.

3 実験

提案モデルが柔軟な予測間隔で次のシーンを予測できるかどうかを確認する実験 (実験 1) と，実際の脳活動データとモデルの隠れた状態の値との相関を用いてモデルの予測機能についての調査 (実験 2) を行い，提案モデルの有効性の評価を行った．また，実験 1 では KITTI Benchmark データセット [2] という車載用データセットを用い，実験 2 では，fMRI で記録したヒトの脳活動データとデータ取得時に視聴した動画像データセットを用いた．動画像データセットの前処理として，動画から静止画を 10fps で抽出し，画像サイズを 160 × 120 ピクセルにダウンサンプリングした．各モデルの Representation 層における特徴表現と脳活動との対応関係を確認するため，リッジ回帰による推定を行い，推定した特徴表現 {R0,R1,R2,R3} と各モデルに刺激動画像を適用して得られた特徴表現との相関係数を求めた．各特徴表現の次元数は，最下層 R0 から R3 の順に，57,600，230,400，115,200，57,600 であった．R1 は内部状態が他層と比べても高次元であり，リソースの制約から推定を行わなかった．脳活動データは，fMRI で観測された 96 × 96 × 72 の全ボクセルのうち，大脳皮質部分に相当する 65,665 次

A Study on a Deep Generative Model Imitating Predictive Coding
†Eri KURODA(kuroda.eri@is.ocha.ac.jp)
‡Shinji NISHIMOTO(nishimoto@nict.go.jp)
‡Satoshi NISHIDA(s-nishida@nict.go.jp)
†Ichiro KOBAYASHI(koba@is.ocha.ac.jp)

図 2: 時間間隔 1 秒での予測.

表 1: 相関係数.

	PredNet			TD-VAE			提案モデル		
α	0.5	1K	25K	0.5	1K	25K	0.5	1K	25K
R0	0.2623	0.2971	0.3207	0.2636	0.2983	0.3285	0.2637	0.2983	0.3291
R2	0.0925	0.1459	0.1955	-	-	-	0.0003	0.0012	0.0016
R3	0.0254	0.1217	0.1871	-	-	-	0.0004	0.0009	0.0012

元のデータを対象とした．各特徴表現と脳活動データのペアを，学習データ 4,497 組，評価データ 300 組として学習を行い，得られたモデルを相関係数を用いて評価を行った.

3.1 実験１：任意の時間間隔での予測

予測間隔の柔軟性について，PredNet と提案モデルでの比較を行った．予測の際の時間幅は 1 秒とし，深層学習のフレームワークは PyTorch を用いて，学習のパラメータは先行研究 [3][4] の設定に基づいた．実験 1 の結果を図 2 に示す．実験結果より，PredNet は時間幅を長くすると出力画像にズレが生じるが，提案モデルは比較的実画像に近い予測画像を生成することが確認できた.

3.2 実験２：各モデルに対する脳活動データとの相関関係

fMRI データとモデルの隠れた状態の値との相関を観測し，モデルの予測に対する性能を調査した．予測タスク中に取得した内部表現の実際の値と，リッジ回帰を用いて脳活動から推定した特徴表現 R0,R2,R3 の対応する値との相関係数を表 1 に示す．リッジ回帰における正則化項の重みである α として，$\{0.50, 1.0 \times 10^3, 2.5 \times 10^4\}$ の範囲の値を試した．また，TD-VAE は階層的な構造を持たないモデルのため，最下層 R0 のみリッジ回帰を行った.

3.3 考察

先行研究 [3][4] を含む全てのモデルに対して，脳活動データから推定した特徴表現 R0 と実際の R0 の相関係数は，α が 2.5×10^4 のとき，約 0.32 であり、これは有意な相関関係を示していると言える．また，PredNet と提案モデルを比較すると，R2 と R3 には相関が見られないが，R0 の相関係数については，提案モデルは PredNet よりもわずかに高いことが確認できた．一方で，TD-VAE と提案モデルの相関係数を比較すると，その差は小さいことが見受けられる．これは，TD-VAE と同様に R0 のみで推論を行うためであり，R2 および R3 の特徴表現の相関は確認することができなかった.

4 おわりに

本研究では，ヒト脳内における予測の仕組みを深層学習モデルとして表現したモデルに対して任意の時間間隔での予測を行う実験と，脳活動データとの相関関係を調査する実験を行い，モデルの有効性を評価した.

謝辞

本研究は，科研費 18H05521 の支援を受けた.

参考文献

[1] 黒田彗莉, 小林一郎. 予測を対象とする深層生成学習モデルを用いた実世界理解への取り組み. 第 82 回全国大会講演論文集, 第 2020 巻, pp. 293–294, feb 2020.

[2] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[3] William Lotter, Gabriel Kreiman, and David D. Cox. Deep predictive coding networks for video prediction and unsupervised learning. *CoRR*, Vol. abs/1605.08104, , 2016.

[4] Karol Gregor and Frederic Besse. Temporal difference variational auto-encoder. *CoRR*, Vol. abs/1806.03107, , 2018.