ヒト脳内での予測を対象とする深層生成学習モデル構築への取り組み

黒田　彗莉　(指導教員：小林 一郎)

1　はじめに

近年，機械学習や深層学習を用いた研究が広く行われ，我々の生活にもその技術の一部が浸透している．それらの技術は，ヒト脳内の意味表象を解読する手法の開発などにも盛んに用いられ，Brain Machine Interface (BMI) 開発における研究の基盤技術の開発や深層学習モデルを作業モデルとしたヒト脳の機能解明などの研究も進められている．しかし，ヒトが時間を認識することに着目した機械学習モデルについては未だ研究はそれほど進んでいるとは言えない．本研究では，大脳皮質における予測メカニズムを模倣した深層学習モデル [1] と様々な時間間隔での将来の予測が可能である深層学習モデル [2] を用いて，前者のモデルでは不規則な時間幅での予測が不可能であったものを様々な時間幅で予測を行えるヒト脳内の情報処理機構を模倣した予測を行う深層生成モデルを構築する．さらに，実験を通して構築したモデルの有効性について検証する．

2　PredNet

PredNet [1] とはヒト脳内の大脳皮質における予測符号化の処理を模倣した機械学習モデルであり，動画像を与えられた元で次の画像を予測するモデルとして提案されている．予測符号化とは，ヒト脳内の大脳皮質で行われているとされる処理である．PredNet では画像の特徴を捉えるのに適した Convolutional Neural Network(CNN)，系列データを扱うのに適した Recurrent Neural Network (RNN) の一種である Long-Short Term Memory(LSTM) と CNN を結合したモデルである Convolutional LSTM を用いて予測タスクを行っている．また，PredNet は同じ構造を持つモジュールを階層的に表現した形をとっており，4 つのモジュールから構成されている．それぞれのモジュールは，脳内の予測モデルを表現した 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 ユニットになっている．PredNet モデル全体の流れは，上位層で生成された予測が Representaion モジュールを介して下位層へ，下位層で生成されたエラーシグナルが上位層へ伝播をすることで，情報のやり取りが行われている（図 1 参照）．

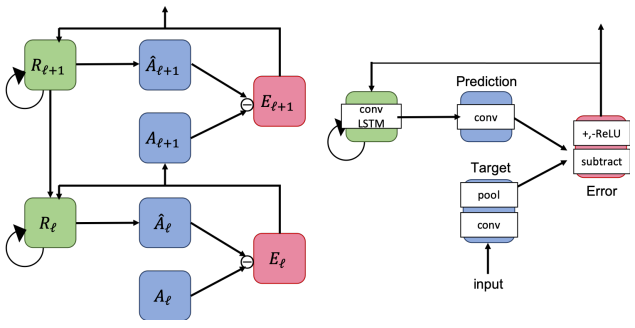


図 1: PredNet 概要図.

3　TD-VAE

Temporal Differential Variational Auto-Encoder (TD-VAE) とは，部分観測値を元に隠れマルコフモデルに基づき状態遷移を行う強化学習モデルである Partially Observable Markov Decision Process (POMDP) に対して信念状態を取り入れた深層生成モデルであり，動画像を与えられた元で自由な時間間隔での予測を可能にしたモデルとして提案されている．TD-VAE では、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) を用いて，予測タスクを行っている．また，TD-VAE は入力情報を観測し，その背景に信念という状態変数を用いて予測対象となるシステムの振る舞いを，第三者となる観測者の視点を取り入れて予測を行う．TD-VAE 全体の流れは，入力情報を観測し，その情報がシステムの挙動が組み込まれた信念層に伝播され，prediction モジュールを介して任意の時間における予測の推論を行う．その推論情報は decoder モジュールを通ることで任意の時間間隔での予測として生成される（図 2 参照）．

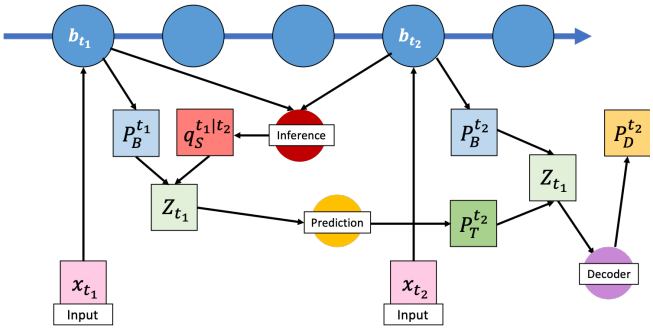


図 2: TD-VAE 概要図.

4　提案モデル

本研究では，上記二つの先行研究 [1][2] を元に，予測を対象とした深層生成モデルの構築を行った．ヒト脳内における情報処理機構を模し，その上で自由な時間幅での予測が可能なモデルの構築を行った．概要を図 3 に示す.

先行研究 PredNet [1] は，ヒト脳内における予測符号化を模倣したモデルになっている．しかし，PredNet における学習は過去の事象に基づいて少し先の将来を予測するものであるため，予測時間幅に対する柔軟性が欠如している．一方，先行研究 TD-VAE [2] は，任意の時間間隔での予測を可能にしたモデルになっている．しかし，ヒトの脳内情報処理機構を反映させることを意図して構築されていない．これらを踏まえた上で，本研究では両者を融合し，ヒトの脳内情報処理機構を模倣した，任意の時間間隔での予測が可能な深層生成モデルの提案を行う．

TD-VAE モデルを元に PredNet モデルを組み込むことで新しく構築したモデル全体の処理の流れについて説明する．現在の時刻を t_1，予測したい未来の時刻を t_2 と仮定する．扱うデータは系列情報であり，信念層と呼ばれる層に観測情報が直接入力される．信念層はモデルのシステムの動きを客観的に見るために定義された層であり，この層には過去から現在までの観測した情報が蓄積されている．時刻 t_1 から時刻 t_2 を予測するために，まず信念層から時刻 t_1 における推論 z_{t_1} が生成される．また，本モデルでは系列状態を扱っているので，信念層は全ての系列情報を知っていることになる．そのため，時刻 t_2 から時刻 t_1 の推論 $\hat{z}_{t_1}$ が生成可能である．このことが PredNet とは異なり，様々な時間幅での予測を可能にしている．推論 z_{t_1} と推論 $\hat{z}_{t_1}$ は同じ時刻 t_1 の状態を表現しているので，推論が等しくなる必要がある．学習は双方の推論の誤差が小さくなるように行う．そしてその差分をエラーシグナルとして信念層の上位層に伝播させる．このように誤差を上位層に伝播させる仕組みはヒトの脳内のボトムアップの仕組みを模倣している．そして，時刻 t_1 における推論 z_{t_1} から prediction 層を経て時刻 t_2 における推論 z_{t_2} を生成する．その推論 z_{t_2} が decoder 層を経て出力画像が生成される．

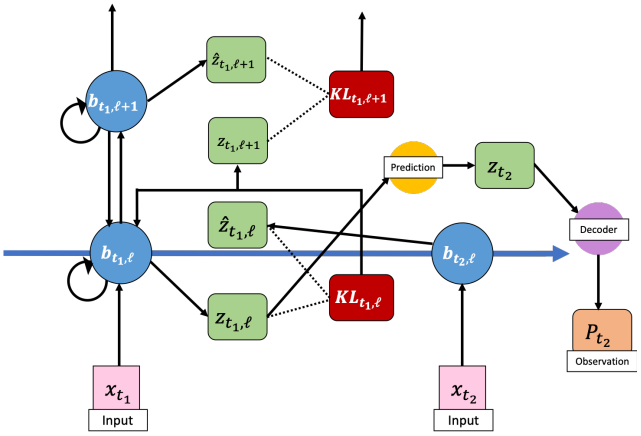


図 3: 提案モデルの概要図．

5 実験

先行研究 [2] と同様の Moving MNIST [3] における学習により，モデルの有効性を確認する (実験 1)．また，提案モデル内の PredNet の枠組みのみを用いた学習を行うことで，本モデル構築の整合性についても検証する (実験 2)．

5.1 データセット

機械学習の動画像データデットである Moving MNIST [3] を用いた．このデータセットは手書きの数字が壁に跳ね返って動いている画像であり，学習用の画像が 6 万枚，テスト用の画像が 1 万枚で構成されている．また各画像にはラベルも付与されている．

5.2 提案モデルにおける実験 (実験 1)

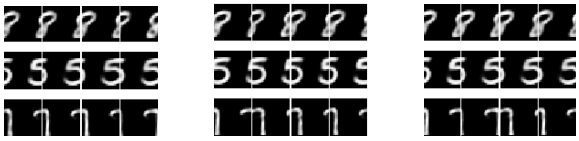
提案モデルについて，動画像データセット Moving MNIST [3] を用いて学習を行った．この実験は先行研究 [2] 内でも任意の時間間隔での予測の可否を調査するため行われている．出力として，任意時間経過した後の数字の挙動についての予測の可否を調査するとともに，モデルが正しい推論を行っているかを確認した．また提案モデルの実装に際しては，深層学習のフレームワーク Pytorch を用いて，学習に関するハイパーパラメータは先行研究 [2] の設定に基づいた．学習は最下層における推論と，任意時刻からの推論の誤差を最小化する形で行った．

5.3 PredNet における実験 (実験 2)

提案モデル内の PredNet の枠組みを用いて，動画像データセット Moving MNIST [3] による学習を行った．また提案モデルにおける PredNet の実装に際しては，深層学習のフレームワーク Pytorch を用いて，学習に関するハイパーパラメータは先行研究 [2] の設定に基づいた．学習は最下層における推論と，任意時刻からの推論の誤差を上位層に伝播させ，誤差を小さくする形で行った．

5.4 結果と考察

実験 1 の結果を図 4 に示す．出力は予測の幅が 1，3，ランダムとなるように生成した．実験結果より，本研究において提案した新しいモデルが深層生成モデルとして正しく挙動したことが確認できた．



(i) 間隔 1　(ii) 間隔 3　(iii) 間隔ランダム

図 4: 任意の時間間隔における出力結果．

実験 2 の結果を図 5 に示す．実画像と予測画像を比較すると，予測が正しくできている数字もあれば，正確ではないものもあることがわかる．このことから PredNet のみの学習についての精度は低いので，さらなる検討が必要だと言える．

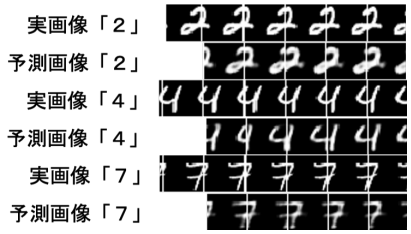


図 5: 提案モデル内の PredNet の枠組みでの出力結果．

6 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と任意の時間間隔での予測を可能にした機械学習モデルを組み合わせ，ヒトの予測を模した新しい深層生成モデルの提案を行った．また，実験を通じて提案モデルの有効性を検証した．

参考文献

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

[2] K. Gregor, G. Papamakarios, F. Besse, L. Buesing, and T. Weber. ”Temporal Difference Variational Auto-Encoder.” In ICLR, 2019.

[3] N. Srivastava, E. Mansimov, and R. Salakhutdinov. ”Unsupervised Learning of Video Representations using LSTMs.” In ICML, 2015.