

# 複雑環境下でのロボット学習に向けた 深層状態空間モデルを用いた映像予測

03-180961 近藤生也  
指導教員 松尾豊 教授

## 1. 序論

近年、機械学習・深層学習を用いてロボットの制御方法をロボット自らに学ばせるロボット学習の研究が進んでおり、ロボットの実用化の可能性が広がっている。ロボット学習において映像予測は重要であり、ロボット自身の行動計画の作成や [1]、ユーザーが事前にロボットの行動を評価する際に用いることができる [2]。映像予測のアプローチは大きく回帰型ニューラルネットワークベースの手法と深層状態空間モデルベースの手法に分けられる。前者は生成精度は高いが長期の予測の際には誤差が蓄積しやすく、また予測を行うには直前までの映像を複数フレーム用意する必要があり予測してから行動したいようなロボット実機の問題設定には向いていない。一方後者は近年深層強化学習の分野で用いられており、ゲームや単純なシミュレーターを題材にした映像予測は可能なものの、複雑なデータに対しては RNN ベースの手法と比較し高精度な予測が難しく、映像予測自体を目的にした研究は進んでいない。

本研究ではロボット学習への応用を前提とし、DSSM をベースにした映像予測性能の向上を目指す。提案手法として、状態表現の階層性に着目した DSSM の拡張を考え、実験によりその有効性を示す。

## 2. 前提知識

### 2.1. DSSM

深層状態空間モデル (Deep State Space Model, DSSM) は将来の観測の予測を通して環境の状態表現を獲得する深層学習手法である [3]。DSSM のグラフィカルモデルは図 1 のように表され、各時刻の状態表現  $s_t$  の遷移とその時刻の観測  $o_t$  の生成過程をニューラルネットワークで表現する。DSSM は初期状態  $s_0$  と環境中の行動主体の行動系列  $a_{1:T}$  を入力とし、 $o_{1:T}$  を予測し出力することができる。ただし初期状態は  $o_0$  から求められる。

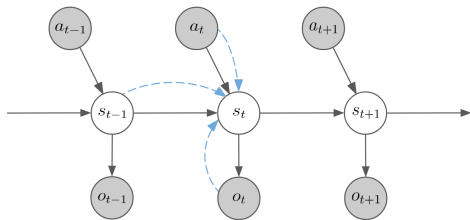


図 1: DSSM のグラフィカルモデル。実線は生成分布、点線は推論分布を表す。簡単のため推論分布は時刻  $t$  でのみ記載している。

DSSM のネットワークのパラメータは、図 1 の点線で表された推論分布  $q(s_t | s_{t-1}, a_t, o_t)$  を導入すること

で、次の変分下限の最大化によって求められる。

$$\log p(o_{1:T} | a_{1:T}) \geq \sum_{t=1}^T (L_{reconst} - L_{KL}) \quad (1)$$

ただし、

$$L_{reconst} = \mathbb{E}_{s_t \sim q(s_t | s_{t-1}, a_t, o_t)} [\log p(o_t | s_t)]$$

$$L_{KL} = \mathbb{E}_{s_{t-1} \sim q(s_{t-1} | s_{t-2}, a_{t-1}, o_{t-1})} [\text{D}_{KL}(q(s_t | s_{t-1}, a_t, o_t) || p(s_t | s_{t-1}, a_t, o_t, s_t))]$$

ここで  $-L_{reconst}$  は観測の再構成誤差を、 $L_{KL}$  は  $s_t$  の生成分布と近似分布の間のカルバック・ライブラー情報量を指す。

## 3. 問題設定

本研究では行動条件付き映像予測の問題を扱う。具体的には、行動系列と観測系列の組  $\{a, o\}$  の訓練用のデータを用いて学習し、評価用のデータで初期観測  $o_0$  と行動系列  $a_{1:10}$  が与えられたときに正しく  $o_{1:10}$  の予測・生成ができることを目指す。本研究では BAIR Push Dataset [4] を用いて各手法の評価を行う。

## 4. 深層状態空間モデルの問題点

予備実験として DSSM を用いて BAIR Push Dataset で学習を行った際の学習曲線を図 3 の Baseline (64) ~ (1024) に示す。ただし括弧内の数字は状態変数  $s_t$  の次元数である。モデルが状態表現として獲得できる情報量を増やすためには素朴には状態変数の次元を大きくすることが考えるが、実際には学習がうまく進まず、性能はむしろ下がることがわかった。これを受けて潜在変数の次元を大きくした際にも学習をうまく進める方法を次のように提案する。

## 5. 状態表現の階層性を考慮することによる 深層状態空間モデルの拡張

状態変数の次元を大きくすると、状態変数の遷移の部分で高次元ベクトルから高次元ベクトルへの変換を学習する必要が生じて全体の学習が困難になる。そこで高次元の状態表現の学習時に小さい状態変数で学習した各時刻の状態表現を補助的に用いることを考え、図 2 のグラフィカルモデルで表される DSSM の拡張モデルを提案する。またこのモデルの学習アルゴリズムをアルゴリズム 1 に示す。

## 6. 実験

### 6.1. 実験内容

提案手法とベースラインの比較実験を行う。提案手法は 64 次元と 512 次元の二階層の状態ベクトルを持つモデルと、64 次元と 512 次元と 1024 次元の三階層の状態ベクトルを持つモデルで実験を行った。ベースラインと提案手法の実装の差は必要最小限にとどめ、どちらも学習時には 10 フレーム先までの予測を行った。

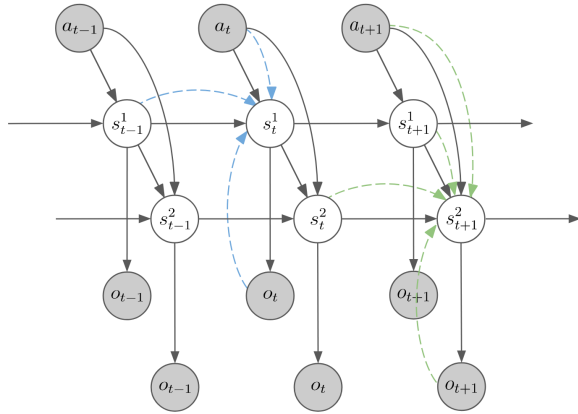


図 2: 提案手法のグラフィカルモデル。実線が生成分布、点線が推論分布を示す。  $s^1, s^2$  の推論分布は簡単のため時刻  $t, t-1$  のみ記載している。また 2 つずつ記載されている  $o_t$  は同じデータを示すが、異なる  $s$  から独立に生成されることを明示している。

#### Algorithm 1 N 階層 DSSM の学習アルゴリズム

**Require:** 階層数  $N$

**for**  $i = 1$  to  $N$  **do**

**while**  $i$  階層の学習が収束していない **do**

$1 \sim i-1$  階層のパラメータを固定し、

$i$  階層を次の目的関数  $L$  で学習する

$$L(a_{1:T}, o_{1:T}) = \sum_{t=1}^T (L_{reconst}^i - L_{KL}^i)$$

        where,

$$L_{reconst}^i = \mathbb{E}_{s_t^i} [\log p(o_t | s_t^i)]$$

$$L_{KL}^i = \mathbb{E}_{s_{t-1}^i} [\mathbb{D}_{KL}(q(s_t^i | s_{t-1}^i, a_t, o_t) \| p(s_t^i | s_{t-1}^i, a_t, o_t, s_{t-1}^{i-1}))]$$

**end while**

**end for**

## 6.2. 実験結果

表 1: 手法ごとの定量評価指標 (尤度)

| 手法                     | 負の対数尤度                                 |
|------------------------|--|
| ベースライン (64)            | $1.1384 \times 10^5$                   |
| 提案手法 (64 + 512)        | <b><math>1.1381 \times 10^5</math></b> |
| 提案手法 (64 + 512 + 1024) | $1.1383 \times 10^5$                   |

表 (1) に最終的な負の対数尤度を示す。

## 7. 考察

課題メイン

## 8. 結論

本論文では、実機ロボットへの応用を見据えて深層状態空間モデルを用いた映像予測について取り上げた。まず深層強化学習などで用いられているシンプルな深層状態空間モデルでは複雑な環境を扱う問題設定に上手くスケールしない問題を示し、その上で深層状態空間モデルの状態表現の階層性を明示的にモデル化した提案手法によってより高次元の状態表現を扱えるよう

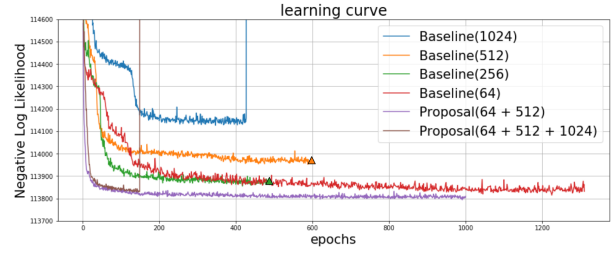


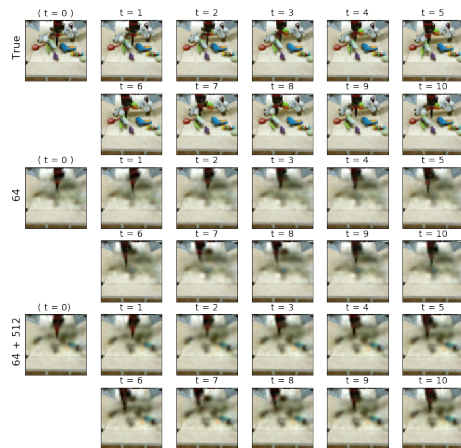
図 3: ベースラインと提案手法の学習曲線。横軸が epoch 数で縦軸が目的関数の値である。グラフ中に三角で示されるのは目的関数の値に Nan が出力されてしまったことを示す。Baseline(1024) は 400epoch を過ぎたあたりで目的関数の値が発散し異常な値をとっている。

にし、さらに映像予測の性能が向上することを示した。実験では、定性的な大きな優位性は示せなかったものの、高次元の状態変数を用いた学習を可能にしたことは深層状態空間モデルの大きな問題を克服したと言える。これにより、これまで映像予測の分野では実機ロボットへの応用上の制約が多いにも関わらず自己回帰的なモデルの研究が主流であったが、状態空間モデルベースの手法が見直されるきっかけになるかもしれない。

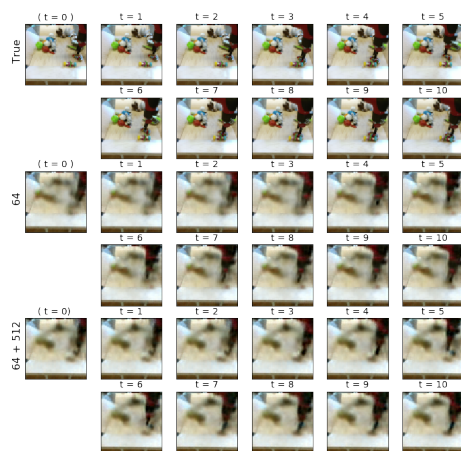
第六章では展望として深層状態空間モデルの研究の方向性を複数上げたが、これらの研究をすすめることによってより高性能な予測が可能になり、また実機への応用の可能性も高められると考える。さらに社会応用の例として、映像予測の実機応用に加え、新たな物理シミュレーションの近似アプローチと新たなロボット学習のあり方の可能性について述べた。この二つの応用例は現段階では可能性の話に過ぎず実現可能かは定かでないがどちらも実現すれば社会的な価値は大きいと考えられるので、今後も慎重に研究を継続していきたい。

## 参考文献

- [1] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning*, pages 2555–2565, 2019.
- [2] Frederik Ebert, Chelsea Finn, Sudeep Dasari, Annie Xie, Alex Lee, and Sergey Levine. Visual foresight: Model-based deep reinforcement learning for vision-based robotic control, 2018.
- [3] Rahul G Krishnan, Uri Shalit, and David Sontag. Deep kalman filters.(2015). *arXiv preprint arXiv:1511.05121*, 2015.
- [4] Frederik Ebert, Chelsea Finn, Alex X. Lee, and Sergey Levine. Self-supervised visual planning with temporal skip connections, 2017.



(a) 改善が見られた例



(b) 改善が見られなかった例

図 4: 映像予測結果 1. 「True」が正解映像, 「64」がベースライン (64) での予測結果, 「64 + 512」が提案手法 (64 + 512) での予測結果を示す. また  $t = 0$  は初期状態の推論時に与えられるフレームを示す.