

平成 31 年度  
卒業論文  
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

平成 32 年 2 月  
指導教員 松尾豊教授

東京大学工学部システム創成学科  
知能社会システムコース  
03-180961 近藤生也

# 目次

第 1 章	序論	1
1.1	本研究の背景 . . . . .	1
1.1.1	行動条件付き映像予測 . . . . .	1
1.1.2	既存の行動条件付き映像予測手法の問題点 . . . . .	2
1.2	本研究の目的 . . . . .	3
1.3	本論文の構成 . . . . .	3
第 2 章	前提知識	4
2.1	変分自己符号化器 (VAE) . . . . .	4
2.2	深層状態空間モデル (DSSM) . . . . .	8
第 3 章	深層状態空間モデルの限界	13
3.1	学習が失敗した例 . . . . .	14
3.2	学習が難しくなる理由の考察 . . . . .	15
第 4 章	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	16
4.1	問題設定の整理 . . . . .	16
4.2	提案手法 . . . . .	16
4.2.1	状態表現の階層性 . . . . .	17
4.2.2	階層的な状態表現の遷移 . . . . .	17
4.2.3	確率モデル・最適化 . . . . .	18
4.3	類似手法との差分 TODO . . . . .	19
第 5 章	実験	21
5.1	実験内容 . . . . .	21
5.1.1	データセット . . . . .	21

5.1.2	ベースラインの実装	22
5.1.3	提案手法の実装	23
5.2	実験結果	24
5.2.1	定量評価(尤度)	24
5.2.2	定性評価	25
<b>第6章 考察</b>		26
6.1	本研究の貢献	26
6.2	課題	26
6.2.1	定性的な改善	26
6.2.2	初期状態の推論	27
6.2.3	低階層の状態表現の再学習	27
6.3	展望	27
6.3.1	敵対的学习の導入	27
6.3.2	多視点からの映像予測	28
6.3.3	他の構造を持つ潜在表現との併用	28
6.3.4	メタ学習の導入	29
6.3.5	映像予測用データ収集の人による代替	29
6.4	社会応用	30
6.4.1	実機ロボットへの応用	30
6.4.2	物理シミュレーションの近似	30
6.4.3	微分可能な環境モデルとしての利用	31
<b>第7章 結論</b>		33
7.1	TODO: 謝辞	34
7.2	その他書いたほうがいいこと	34
<b>謝辞</b>		35
<b>参考文献</b>		36

# 図目次

2.1	VAE のグラフィカルモデル . . . . .	4
2.2	推論分布を導入した VAE のグラフィカルモデル . . . . .	5
2.3	VAE を用いて生成された画像の例 . . . . .	7
2.4	SSM のグラフィカルモデル (TODO: 書き直す) . . . . .	8
2.5	DSSM で生成される映像の例 (回転する数字画像) . . . . .	11
2.6	DSSM で生成される映像の例 (強化学習エージェント) . . . . .	12
3.1	状態変数の次元を変えた時の DSSM の学習曲線 . . . . .	13
3.2	(a) DSSM の学習がうまくいっている例 . . . . .	14
3.3	(b) DSSM の学習が失敗した例 . . . . .	14
3.4	DSSM の学習中にサンプルされた映像の例 . . . . .	14
4.1	hierarchical . . . . .	17
4.2	transition base . . . . .	18
4.3	transition proposal . . . . .	18
4.4	提案手法のグラフィカルモデル . . . . .	18
5.1	提案手法の学習曲線 . . . . .	24
6.1	TODO:この図きれいにする . . . . .	31

# 表目次

# 第1章

## 序論

### 1.1 本研究の背景

#### 1.1.1 行動条件付き映像予測

多様な環境で様々なタスクが遂行可能な汎用的なロボット (generalist robots) の開発はロボット工学の最重要課題の一つである。ロボットハードウェアの低価格化、汎用的なロボットソフトウェアの普及に加え、近年の急速な深層学習技術の発展を受けてロボットの制御方策を自ら学習させるロボット学習の研究が進んでおり、ロボットで遂行可能なタスクは着実に増えている。

ロボット学習において、将来予測、特に映像予測を明示的に学習することは、

- ロボット自身が映像予測を用いた方策をたてることが可能になる
- 映像予測結果を人が評価することでロボットの行動を予め評価できる

という大きく二つの点からで重要であると言える。一点目の映像予測を用いた方策の例として、Hafner ら [1] は、強化学習の問題設定において明示的に学習した映像予測モデルを用いることで行動系列をランダムにサンプリングして評価するような簡単なアルゴリズムで効率的なプランニングが可能であることを示した。二点目の映像予測を人が評価する例として Ebert らによる研究 [2] では学習した映像予測モデルを用いて、ロボットの操作によって予想される物体の移動の軌跡を確率分布として出力することができ、これを用いて人がロボットの行動の正しさを予め判断することができる。

このようにロボット学習における映像予測は重要であるが、映像予測だけを切り取って研究されることも多い。映像予測の中でも、ロボットの行動の結果として観測される映像を予測す

る問題設定を行動条件付き映像予測と呼び、様々な研究がなされてきている。近年高精度な行動条件付き映像予測手法がいくつか提案されており、ロボット学習研究で扱うタスクの高度化を背景にしてこれらの映像予測手法をより複雑な問題設定に対して適用していきたいと考えられているが、いくつかの研究で既存の行動条件付き映像予測は上手く機能しない可能性があることがわかってきていている。ただしここでいう「より複雑な問題設定」とは具体的には環境中に複数の操作対象の物体が隣接し合って置かれている場合、操作対象が布などの非剛体物である場合など、観測の時間変化に多数のパラメータが関与していたり、複雑な物理法則がはたらいているような問題設定を想定している。次に既存の行動条件付き映像予測手法の問題点について示す。

### 1.1.2 既存の行動条件付き映像予測手法の問題点

行動条件付き映像予測手法は大きく再帰型ニューラルネットワーク (RNN) ベースの手法と深層状態空間モデル (DSSM) ベースの手法に分けられる。Hafner ら [1] の研究など深層強化学習の問題で映像予測を明示的に行う場合は後者の DSSM ベースの手法が多く採用されるが、映像予測の問題では前者の RNN ベースの手法が多く使われている [3][4]。

RNN ベースの手法は予測した 1 ステップ先の画像を入力にして更に 1 ステップ先の画像を出力するというような、自らの出力を逐次入力する構造を持つ。RNN ベースの手法は DSSM と比較して高精度な映像を生成に長けている反面、近年提案されている RNN ベースの手法には以下のような問題点がある。

- 「確率的な遷移を考慮できない」、は確率的な遷移 + 自己回帰な論文もあるので入れませんでした
- 誤差が蓄積しやすい
- 文脈を必要とする

一点目について、RNN ベースのモデルを用いると常に直前のフレームを参照して次のフレームを予測するため短い期間の予測であれば精度は高くなるが、予測誤差が蓄積していくために長期の予測には向かないことが示されている [1]。

二点目について、RNN はモデルの内部状態を十分に更新した後でないと適切に予測が行えず、予測を始める前に文脈としてそれより前の数フレームを与える必要があり、この文脈として与えられるフレーム数が少ないと予測が悪化することが知られている [4]。ロボット実機への応用を考えた場合、現在の状態から未来を予測する際に文脈を得るために先に数ステップ行

動することは、予測してから行動するという目的意識に反しており実用的できない。このため RNN ベースの手法をロボット実機に応用する際には、文脈として現在の観測という 1 フレームのみ与れば十分機能するように改善する必要がある。このように、RNN ベースの手法は制約がありそもそも実ロボットへの応用に向いていない可能性がある。

一方、DSSM は各時刻の状態をベクトル（状態ベクトル）で表現し、毎時刻ロボットの行動によって状態ベクトルが遷移し、その時刻に観測される画像は状態ベクトルからの写像であると考えて遷移モデルと写像のモデルを学習する。DSSM は強化学習の分野で長期の予測にも用いられているなど安定した未来の予測に長けているが RNN と比較して画像の生成時に直前の画像を用いないことから高精度な生成は難しく、また映像生成自体を目的にして DSSM を用いた研究は現状少ない。

## 1.2 本研究の目的

これらの研究背景を踏まえ、行動条件付き映像予測の問題をより複雑な問題設定にスケールさせることを目指し、本研究では特に実ロボットへの応用を重要視して DSSM ベースの行動条件付き映像予測に取り組む。まず DSSM で高精度な生成が難しいことを確認しその理由を簡単に考察する。その上で特に複雑な問題設定に広く取り入れることが可能な帰納バイアスを提案しモデルに組み込むことで DSSM を拡張手法を提案する。さらに行動条件付き映像予測用のデータセットを用いて提案手法の有効性についての定性的・定量的な評価を行い、DSSM を使った際にもより高精度な映像予測を可能にすることを目指す。

最後に実験結果を踏まえて、今後の課題と社会応用について述べる。

## 1.3 本論文の構成

本論文の構成は以下の通りである。

第二章では、本研究で中心的に扱う深層状態空間モデル等について説明する

第三章では、深層状態空間モデルの限界を実験的に示し、問題点を指摘する。

第四章では、前章の議論を踏まえて深層状態空間モデルを拡張する提案手法を説明する。

第五章では、実験を行い提案手法の有効性を示す。

第六章では、前章までの議論を踏まえて考察、そして社会応用の可能性について述べる。

最後に第七章で結論を述べる。

## 第 2 章

# 前提知識

本章では、まず深層状態空間モデル (Deep State Space Model, 以下 DSSM) のベースとなる変分自己符号化器 (Variational Auto Encoder, 以下 VAE) について説明し、続いて DSSM の説明を行う。

### 2.1 変分自己符号化器 (VAE)

変分自己符号化器 (VAE) は深層生成モデルの一種である。VAE では、高次元のデータ  $\mathbf{x} \in \mathbb{R}^n$  の背後に比較的低次元の潜在表現  $\mathbf{z} \in \mathbb{R}^m$  があると考え、Fig. 2.1 のようなグラフィカルモデルに従ってデータの分布  $p(\mathbf{x})$  を次式で表す。

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z} \quad (2.1)$$

Fig. 2.1 のグラフィカルモデルは、高次元なデータ  $\mathbf{x}$  は  $\mathbf{x}$  の空間上の非常に限られた領域に局所的に存在しているため、それらを低次元の空間で表現可能であるとする多様体仮設に基づいている。正しい  $p(\mathbf{x}|\mathbf{z})$  のモデルが得られれば、 $\mathbf{z}$  が与えられたときに  $p(\mathbf{x}|\mathbf{z})$  のモデルを



Fig. 2.1 VAE のグラフィカルモデル



Fig. 2.2 推論分布を導入した VAE 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.

使ってそれに対応する  $x$  を生成することができ, また正しい  $p(z|x)$  のモデルが得られると  $x$  のコンパクトな表現としての  $z$  が推論できる. VAE は画像のエンコードやデコードに優れており, 深層強化学習や異常検知など様々な分野に応用されている.

ここからは VAE の理論的な説明を行う. まず VAE では, 以下の 2 つの仮定を置く.

$$p(z) = \mathcal{N}(z|0, \mathbf{I}) \quad (2.2)$$

$$p(z|x) = \mathcal{N}(z|\mu(x), \sigma(x)) \quad (2.3)$$

式 (2.2) は, 潜在表現の空間が標準正規分布に従うという仮定であり, 式 (2.3) は,  $x$  に条件づけられた潜在変数の分布も正規分布に従うという仮定となっている.

VAE ではさらに  $x$  が与えられたときの  $z$  の条件付き確率  $p(z|x)$  を近似する  $q(z|x)$  を導入する. この  $q(z|x)$  は推論分布, 近似分布などと呼ばれ, グラフィカルモデル中に記述すると Fig. 2.2 となる. ただし  $p(z|x)$  を直接考えるのは,  $p(x|z)$  を表すニューラルネットのパラメータで解析的に正しい  $p(z|x)$  を表すことが難しいためである. この  $z$  の推論モデル  $q(z|x)$  と  $x$  の生成モデル  $p(x|z)$  を適当なニューラルネットワークを使ってモデル化すると, それらのパラメータは  $x$  のデータ集合が与えられた際に最尤推定によって求めることができる.

最尤推定の際には式 (2.2) の対数尤度をとって導出される, 以下の変分下限を用いる.

$$\log p(\mathbf{x}) = \log \int p(\mathbf{x}|z)p(z)dz \quad (2.4)$$

$$= \log \int q(z|\mathbf{x}) \frac{p(\mathbf{x}|z)p(z)}{q(z|\mathbf{x})} dz \quad (2.5)$$

$$\geq \int q(z|\mathbf{x}) \log \frac{p(\mathbf{x}|z)p(z)}{q(z|\mathbf{x})} dz \quad (2.6)$$

$$= \int q(z|\mathbf{x}) \log p(\mathbf{x}|z) dz - \int q(z|\mathbf{x}) \log \frac{q(z|\mathbf{x})}{p(z)} dz \\ = \mathbb{E}_{z \sim q(z|\mathbf{x})} [\log p(\mathbf{x}|z)] - D_{KL}(q(z|\mathbf{x}) \| p(z)) \quad (2.7)$$

式 (2.6) にはイエンセンの不等式を用いている。式 (2.5) では、数値計算をする都合上式の  $z$  の周辺化が難しいため、先述した  $q(z|\mathbf{x})$  を導入し、更に式 (2.7) 第一項で  $q(z|\mathbf{x})$  からサンプリングされる  $L$  個の  $z$  を用いて  $\frac{1}{L} \sum_l \log p(\mathbf{x}|z)$  でモンテカルロ近似することによって、周辺化を排除している。ただし通常  $L = 1$  で計算される。式 (2.7) の第 2 項の  $D_{KL}$  (カルバックリブラー距離) は、いま  $p(z)$ ,  $q(z|\mathbf{x})$  共にガウス分布を仮定しているため解析的に計算することができ、また式 (2.7) の第 1 項は、 $p(\mathbf{x}|z)$  の分布に正規分布やベルヌーイ分布を仮定することで二乗誤差やクロスエントロピー誤差として計算できる。この変分下限を目的関数にすることで生成モデル  $p(\mathbf{x}|z)$  と推論モデル  $q(z|\mathbf{x})$  を同時に学習することができる。

以上が VAE の概要である。このように学習された VAE は、事前分布  $p(z)$  から  $z$  をサンプリングし、デコーダを通すことで Fig. 5.1 のような新たなデータを生成することができる。左は顔画像のデータセット Frey Face を用いて学習した VAE によって新たに生成された画像、右は MNIST で学習した VAE によって新たに生成された画像である。この 2 つの図は、それぞれ 2 次元の潜在変数をおいた VAE で学習したのち  $z$  の値を少しずつずらしながら新しい画像を生成したもので、なめらかに変化した画像が生成できていることから画像空間という高次元の空間上のデータを上手く低次元の空間に埋め込んでいることがわかる。

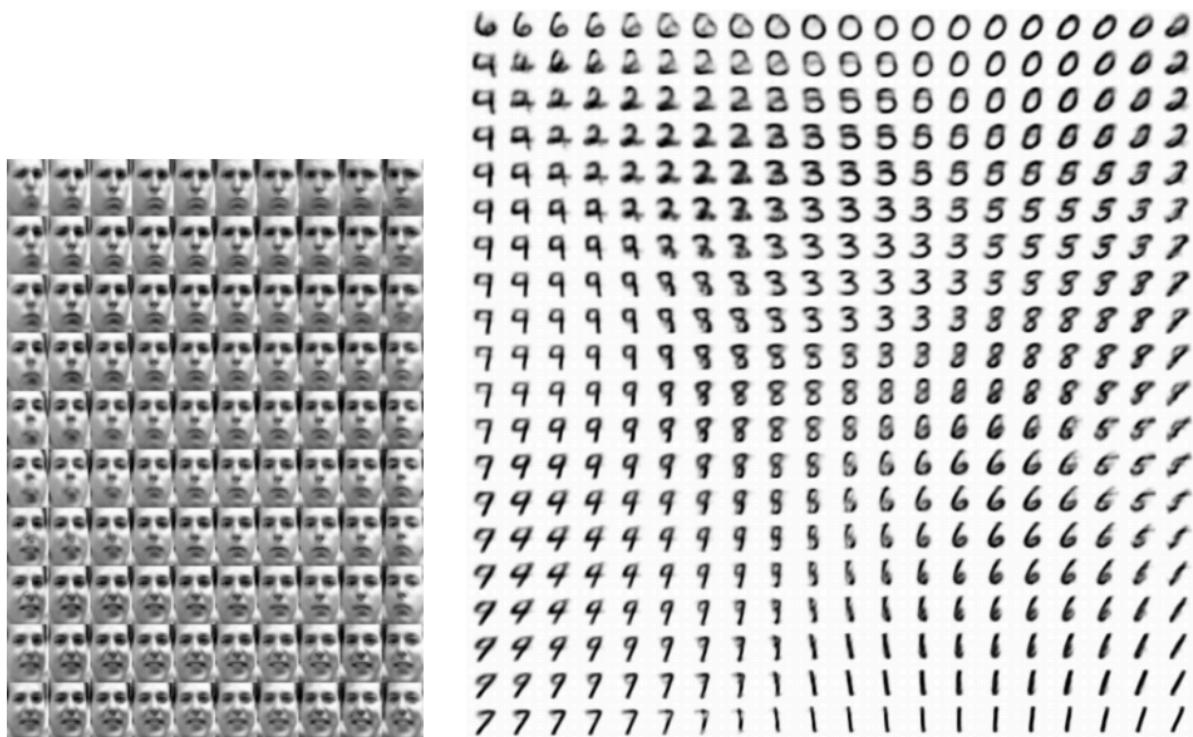


Fig. 2.3 VAE を用いて生成された画像の例 ([5] より引用)

## 2.2 深層状態空間モデル (DSSM)

深層状態空間モデル (Deep State Space Model: DSSM, Deep Karman Filters, Deep Markov Model, 単に SSM などとも呼ばれる) について説明する。VAE がデータ一つ一つの潜在表現を考えるのに対し, DSSM は時間変化があるデータの各時刻の潜在表現を考えて更にその潜在表現の時間変化をモデル化したものであり, VAE を時系列方向に拡張したモデルとみなすこともできる。DSSM ではこの潜在表現のことを状態表現と呼ぶ。

DSSM は特に深層強化学習の分野で用いられており, エージェントがある環境中で行動を起こした結果として視覚フィードバックや報酬フィードバックなどが観測されるときに, DSSM で将来の観測を正しく予測できるよう学習することで良い環境の状態表現を獲得することができ, さらにその状態表現を用いることで良い行動方策が得られる。[引用]

本論文では Fig. 2.4 のようなグラフィカルモデルで表される DSSM のモデルを考えるが, 行動系列が与えられない場合や, 強化学習のように環境の状態に応じて報酬が与えられる問題においても以下の説明は同様にして考えることができる。

ここから DSSM の理論的な説明を行う。DSSM は式 (2.8) 式が示すように, 初期状態  $s_0$

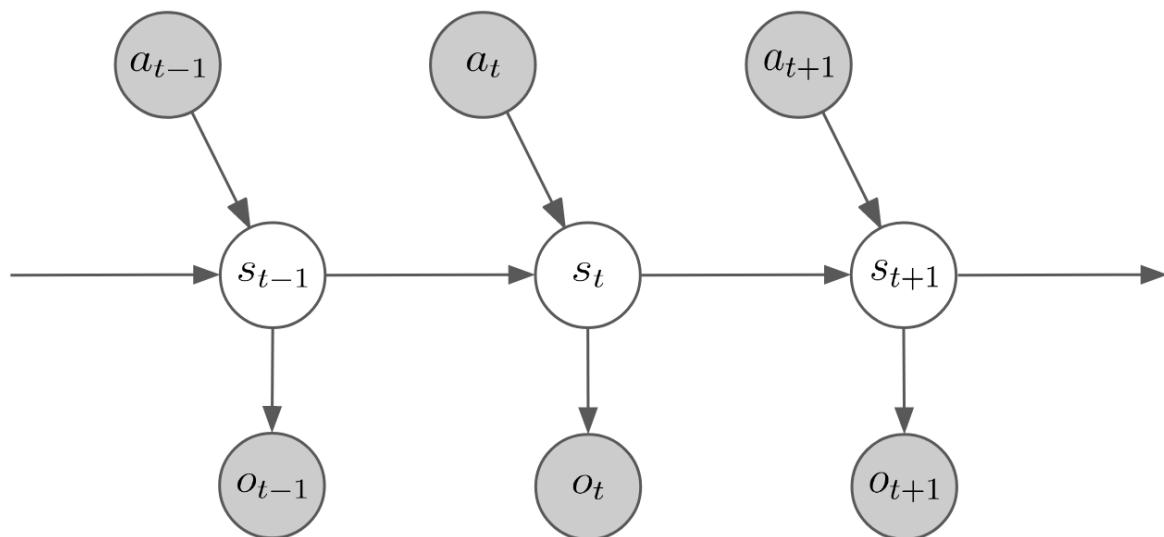


Fig. 2.4 SSM のグラフィカルモデル (TODO: 書き直す)

と行動系列  $a_{1:T}$  で条件付けられた観測  $o_{1:T}$  を予測するモデルである。DSSM では各時刻の観測  $o_t$  は各時刻の状態表現  $s_t$  から生成されると考え、各時刻の状態表現  $s_t$  を明示的に推論する。初期状態  $s_0$  の定め方は任意だが、多くの場合ゼロベクトルが用いられる。以降、簡単のため条件  $s_0$  は省略する。

$$\begin{aligned} p(o_{1:T}|a_{1:T}, s_0) &= \prod_{t=1}^T p(o_t|a_{1:t}, s_0) \\ &= \prod_{t=1}^T \int p(o_t|s_t) p(s_t|s_{t-1}, a_t) ds_t \end{aligned} \quad (2.8)$$

DSSM では以下の仮定をおく。式 (2.9) は各時刻の状態ベクトル  $s_t$  は正規分布に従うという仮定で、式 (2.10) は  $o_t$  で条件付けられた各時刻の状態ベクトル  $s_t$  も正規分布に従うという仮定である。

$$p(s_t|s_{t-1}, a_t) = \mathcal{N}(s_t|\mu(s_{t-1}, a_t), \sigma(s_{t-1}, a_t)) \quad (2.9)$$

$$p(s_t|s_{t-1}, a_t, o_t) = \mathcal{N}(s_t|\mu(s_{t-1}, a_t, o_t), \sigma(s_{t-1}, a_t, o_t)) \quad (2.10)$$

さらに DSSM では VAE と同様に  $s_t$  の推論モデル  $q(s_t|s_{t-1}, a_t, o_t)$  を導入し、生成過程と推論モデルをニューラルネットによってモデル化し、最尤推定によってそれらのパラメータを求める。

以下の変分下限で最尤推定を行う。

$$\begin{aligned}
\log p(o_{1:T}|a_{1:T}) &= \log \prod_{t=1}^T p(o_t|a_{1:t}) \\
&= \log \prod_{t=1}^T \int p(o_t|s_t) p(s_t|s_{t-1}, a_t) ds_t \\
&= \sum_{t=1}^T \log \int p(o_t|s_t) p(s_t|s_{t-1}, a_t) ds_t \\
&= \sum_{t=1}^T \log \int q(s_t|s_{t-1}, a_t, o_t) \frac{p(o_t|s_t)p(s_t|s_{t-1}, a_t)}{q(s_t|s_{t-1}, a_t, o_t)} ds_t
\end{aligned} \tag{2.11}$$

$$\geq \sum_{t=1}^T \int q(s_t|s_{t-1}, a_t, o_t) \log \frac{p(o_t|s_t)p(s_t|s_{t-1}, a_t)}{q(s_t|s_{t-1}, a_t, o_t)} ds_t \tag{2.12}$$

$$\begin{aligned}
&= \sum_{t=1}^T \left( \int q(s_t|s_{t-1}, a_t, o_t) \log p(o_t|s_t) ds_t \right. \\
&\quad \left. - \int q(s_t|s_{t-1}, a_t, o_t) \log \frac{p(s_t|s_{t-1}, a_t)}{q(s_t|s_{t-1}, a_t, o_t)} ds_t \right) \\
&= \sum_{t=1}^T \left( \mathbb{E}_{s_t \sim q(s_t|s_{t-1}, a_t, o_t)} [\log p(o_t|s_t)] \right. \\
&\quad \left. - \mathbb{E}_{s_{t-1} \sim q(s_{t-1}|s_{t-2}, a_{t-1}, o_{t-1})} [\text{D}_{\text{KL}}(q(s_t|s_{t-1}, a_t, o_t) \| p(s_t|s_{t-1}, a_t, o_t))] \right)
\end{aligned} \tag{2.13}$$

式 (2.12) にはイエンセンの不等式を用いている。式 (2.11) では、 $s_t$  の周辺化を排除するために先述した近似分布  $q(s_t|s_{t-1}, a_t, o_t)$  を導入し、さらに式 (2.13) で  $s_t, s_{t-1}$  をサンプリングしてモンテカルロ近似を行う。式 (2.13) の第 2 項のカルバックライブラ一距離は、いま  $p(s_t|s_{t-1}, a_t, o_t)$ ,  $q(s_t|s_{t-1}, a_t, o_t)$  共にガウス分布を仮定しているため、解析的に計算することができ、また式 (2.7) の第 1 項は、 $p(o_t|s_t)$  の分布に正規分布やベルヌーイ分布を仮定することで尤度が計算できる。

この変分下限を目的関数にすることで各モデルを学習することができる。また学習時には 10 フレーム程度先までの予測を行うのが一般的であるが、学習時の予測の長さとモデルの評価時の予測の長さを揃える必要はなく、評価時にはより長期の予測も可能である。

最後に  $s_0$  の求め方について、初期状態が必ず一定な問題設定ではゼロベクトルなど一定の値に決め打ちすることができるが、様々な初期状態が考えられる場合は何らかの方法で推論する必要がある。本研究では慣例に習って  $s_0$  を  $s_0 \sim q(s_0|\vec{0}, \vec{0}, o_0)$  によって定め、(2.13) の第 2 項の  $s_0$  のサンプリングは ( $o_{t=-1}$  や  $a_{t=-1}$  が与えられない前提を考えるため) これで置き換

える。 $s_0$  の推論方法については議論の余地があると考えられ、これについては考察 (TODO リンクする)『 $s_0$  の推論』で触れる。

以上が DSSM の概要である。DSSM を用いると、Fig. 2.5, Fig. 2.6 のように数フレーム先を予測することができる。Fig. 2.5 は、行動系列として回転情報が与えられており、回転する画像の予測を行う。Fig. 2.6 は四足歩行するシミュレータ上のエージェントの第三者視点からの観測を予測しており、行動系列としてエージェントの各関節に加えられる力が与えられている。DSSM は長期の予測をした際にも生成自体が安定している点で自己回帰モデルと比較して優れている。

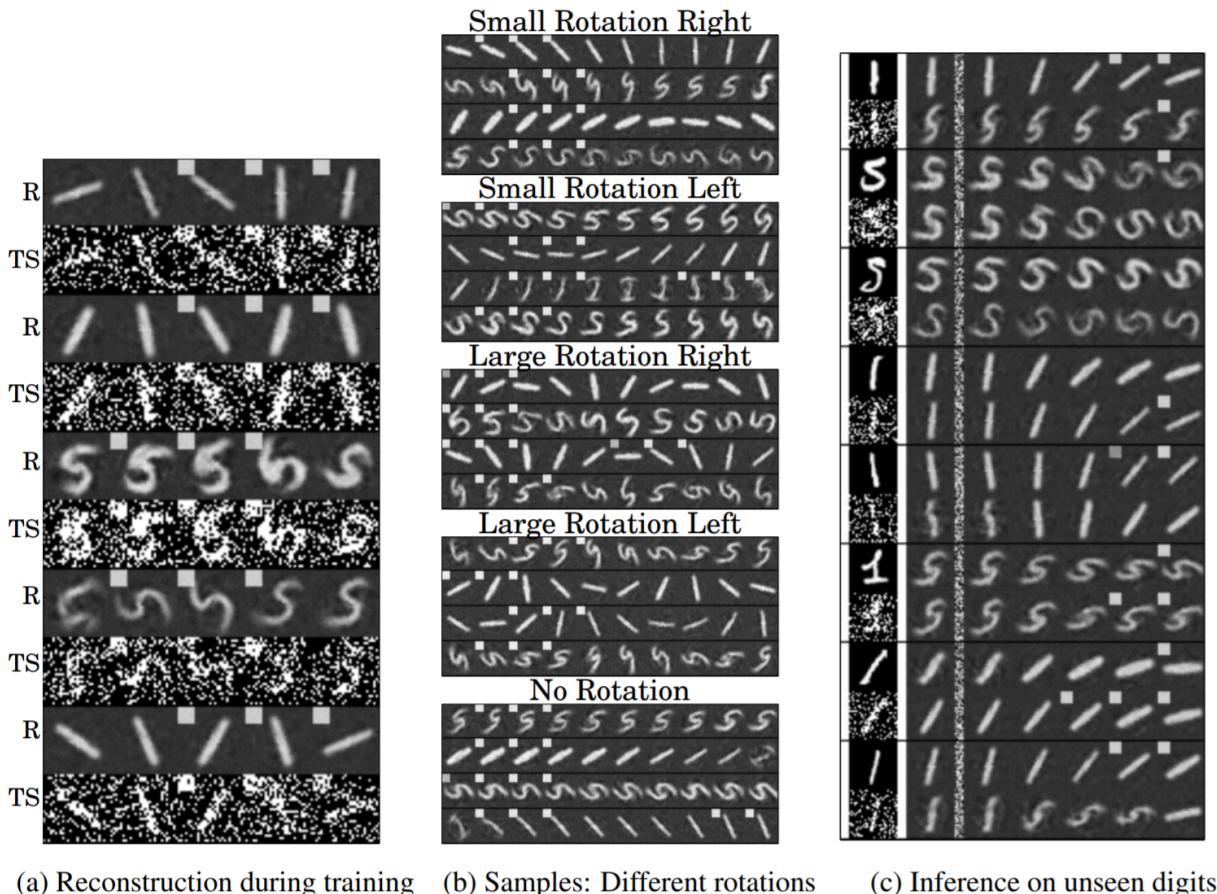


Fig. 2.5 DSSM で生成される映像の例 (回転する数字画像)(TODO: 引用文献入れる、右図 c だけにする)

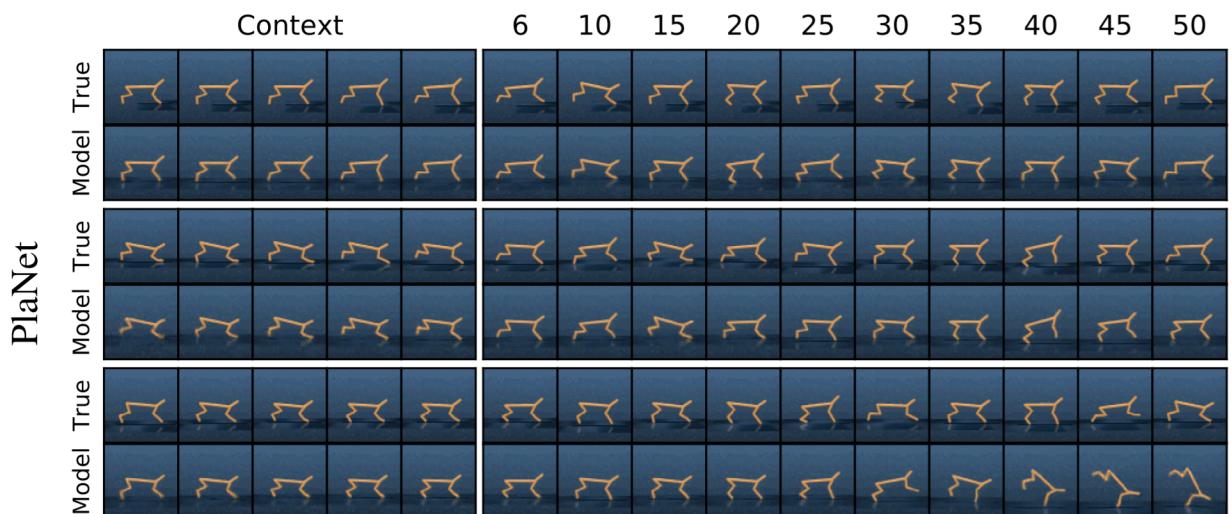


Fig. 2.6 DSSM で生成される映像の例 (強化学習エージェント). ただし SSM を少し拡張した RSSM モデルを使っている

## 第 3 章

# 深層状態空間モデルの限界

深層状態空間モデルを用いてより複雑な環境の映像予測を考える場合、潜在的な情報が多くなるはずであるためより大きな状態表現を扱えるようモデルを大きくする必要がある。素朴には状態変数の次元を大きくすることがよいと考えられるが、予備実験の中で深層状態空間モデルは状態変数の次元を大きくすると学習が困難になる、または学習が安定しにくくなることがわかった。まずこのことを示した上で、この理由を考察する。

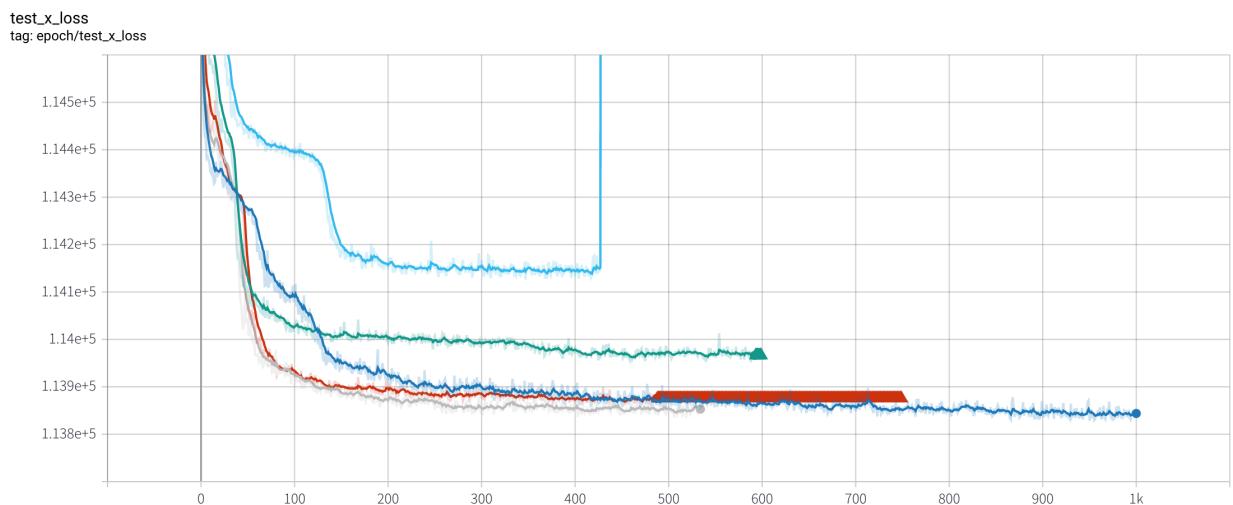


Fig. 3.1 状態変数の次元を変えた時の DSSM の学習曲線。横軸が epoch 数で縦軸が目的関数の値である。グラフ中に三角で示されるのは目的関数の値に Nan が出力されてしまったことを示す。(TODO: 綺麗にする, 凡例入れる)



Fig. 3.2 (a) DSSM の学習がうまくいっている例

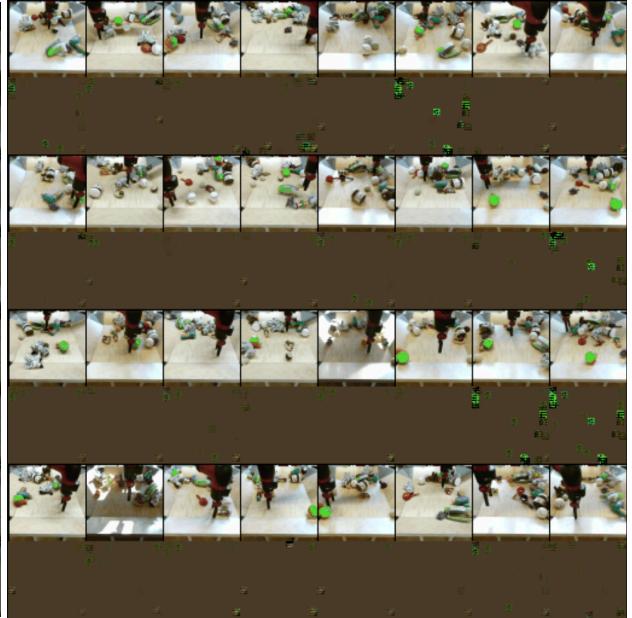


Fig. 3.3 (b) DSSM の学習が失敗した例

Fig. 3.4 DSSM の学習中にサンプルされた映像の例. 上から奇数段目が正解映像で, 上から偶数段目が一段上の映像の行動条件付き予測結果である. 10 フレームの映像予測を行っているが, 図ではある一時刻の観測のみが示されている (TODO: (a) と (b) の Fig. の字消す)

### 3.1 学習が失敗した例

Fig. 3.1 は, 本論文の 4 章以降でベースラインとして用いる通常の DSSM モデルを状態変数の次元を何通りかに変えてモデルを構築し, 学習時の訓練用データでの目的関数の値の増減をプロットしたものである. 状態変数を比較的低次元の 64 次元に設定すると順調に学習が進み, Fig. 3.1 の左図に示すように少しずつ近い映像が出力されるようになる. しかし状態変数の次元を 512 次元に設定すると明らかに精度が悪化し, 1024 次元にするとさらに悪化している. Fig. 3.1 の右図は 1024 次元での実験で目的関数の値が発散した後の生成映像の様子で, 他の次元で学習が失敗した場合も同じような映像が生成される. 256 次元や 512 次元での実験で途中から目的関数の値に Nan が出力されてしまうのは, 5 章の『実験の安定化』(TODO: リンク埋める) で述べるように, カルバックライブラリ距離を小さくすることに失敗しゼロ除算が発生したためである. 1024 次元で値が発散している理由は定かでないが, 256 次元, 512 次元と同様の理由であると考えられる.

ELBO が Nan をとったり発散したりすることは数値計算の丸め誤差などの理由も関わるの

で一旦考えないとしても、明らかに状態変数の次元を大きくすることで精度が下がっていることが Fig. 3.1 からわかる。

## 3.2 学習が難しくなる理由の考察

VAE の学習では前節で述べたような問題は起こらないことを踏まえると、DSSM で学習が難しいのは状態変数の遷移の部分であると考えられる。さらに、状態変数に高次元を仮定したとき、状態変数の遷移の部分では状態変数の生成モデル・推論モデルとともに高次元ベクトルから高次元ベクトルへの写像を学習する必要が生じてまずこの部分で学習に時間がかかりやすくなってしまっており、さらに推論モデルが十分に学習されないまま状態変数の事前分布と事後分布のカルバックライブラ一距離の最小化が図られてしまうので、学習が安定しにくくなっていると考えられる。

このことから、単に状態変数を高次元にすることは DSSM の性質上適切ではなく、DSSM をより複雑な環境を扱う問題にスケールさせるためには他の方法を考える必要がある。以上の予備実験を受け、次章ではシンプルな DSSM の拡張方法を提案する。

## 第4章

# 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

第三章の問題を受け、第四章ではシンプルな帰納バイアスを導入することによって DSSM を拡張する方法を提案する。はじめに本研究で扱う問題設定について改めて整理し、続けて提案手法とその既存の類似手法について述べる。

### 4.1 問題設定の整理

本研究では行動条件付き映像予測の問題を解く。具体的には、ある行動主体が実行した行動系列  $a_{1:10}$  と初期観測  $o_0$  が与えられたときに  $o_{1:10}$  を生成し、その生成される映像の尤度を高めることを目指す。ただし訓練時には行動系列と観測系列の組  $\{\vec{a}, \vec{o}\}$  の訓練用のデータセットを用いることができ、評価時には、訓練データには含まれないが訓練データと同じ条件で収集された評価用のデータセットを用いる。本研究の目的は、この条件付き映像予測問題においてベースラインに DSSM を設定し、DSSM をより複雑な環境での映像予測問題にもスケールできるように拡張することである。

### 4.2 提案手法

三章で述べたようにベースラインの DSSM では潜在変数の次元を大きくすると学習がうまくいかなかった。しかし予備実験で得た、低次元の潜在変数を用いたときには部分的に学習が進んだという事実をヒントにし、状態変数の次元を大きくしていく方向性はそのままで状態表現の階層性を考えることにより、より複雑な問題設定においても学習可能な DSSM の拡張方法を提案する。

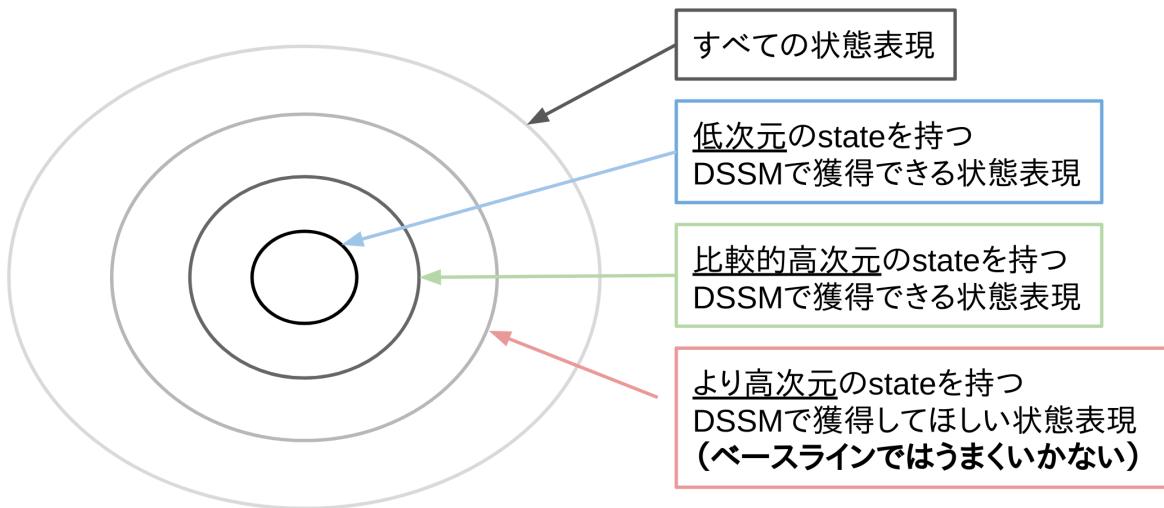


Fig. 4.1 hierarchical

#### 4.2.1 状態表現の階層性

はじめに、ベースラインの DSSM において潜在変数の次元を変えた時に獲得される情報について考察する。低次元の状態変数で獲得できる情報は高次元の状態変数を用いた場合にも当然獲得できると考えた場合、図 (4.1) のように高次元の状態変数が持つ情報は低次元の状態変数が持つ情報をほぼ内包していると考えることができる。ここで状態変数の次元をより大きくしたときに精度がむしろ悪化することが問題であったが、これは三章で述べたとおり、状態変数の次元が大きくなったときに高次元ベクトルから高次元ベクトルへの写像を学習する必要が生じるべき写像先がなかなか定まらないことが原因であると考えられ、何らかの方法で遷移モデルの学習を補助することで図のようにより多くの情報を獲得できる可能性がある。

#### 4.2.2 階層的な状態表現の遷移

ベースラインの状態表現の遷移は図 (4.2) が示すように状態変数が持つすべての情報を一度に変換することを考えているが、直感的に一括で変換することは学習が難しいと思われる。状態表現を一括で変換する代わりに、前節のような階層性の概念を導入することで図 (4.3) のように簡単に遷移が学習できる部分から順に遷移させていくような方法を考えることができる。

このような階層的な状態表現の遷移を考えると、はじめから高次元の状態表現の遷移を考えずに学習の習熟度に合わせて徐々に高次元の状態ベクトルの遷移を学習することができ、学習

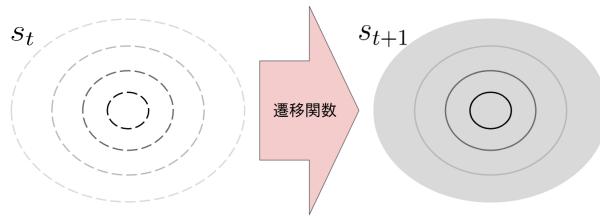


Fig. 4.2 transition base

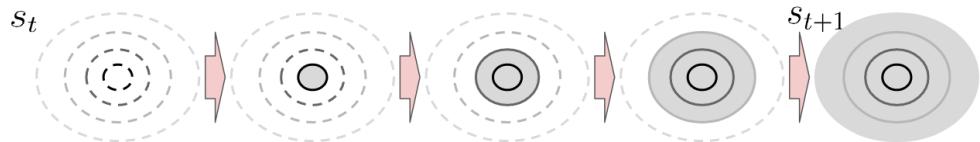
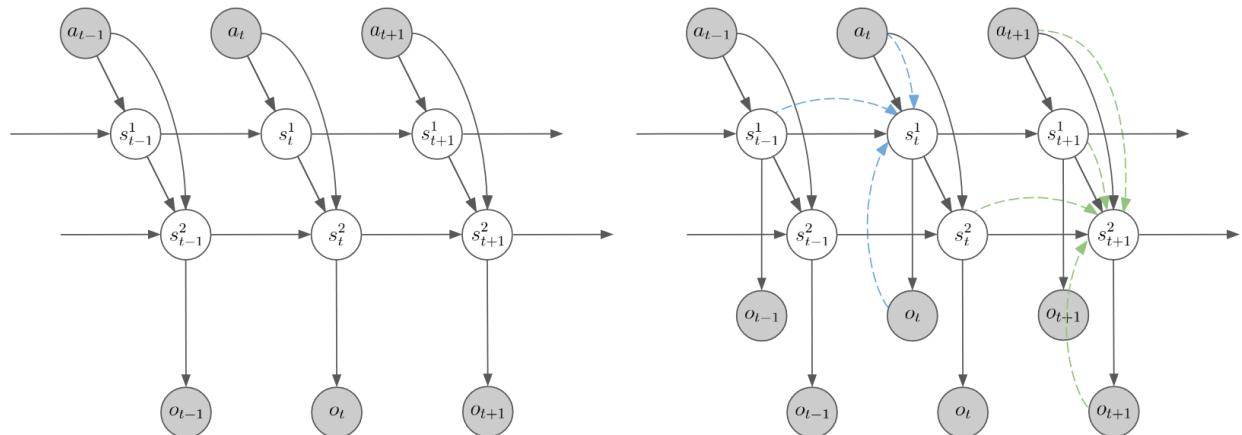


Fig. 4.3 transition proposal

Fig. 4.4 提案手法のグラフィカルモデル. 点線の  $s^1, s^2$  の推論分布は簡単のため時刻  $t$ ,  $t-1$  でのみ記載している. proposal (学習時) 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.

がスムースに進みやすくなると考えられる。

#### 4.2.3 確率モデル・最適化

ここまでで状態変数の階層性とその遷移を考えたが、この階層性の仮定は DSSM の性能の向上に十分寄与しうると考え、状態変数の階層性を帰納バイアスとして DSSM に組み込み以下ののようなモデルとその最適化アルゴリズムを提案する。

---

**Algorithm 1** N 階層 DSSM の学習アルゴリズム TODO: 書き直す

---

```

for 一階層目から N 階層目まで do
  while 学習が収束していない do
    現在の階層より上の階層のパラメータを固定し,
    現在の階層を以下の目的関数で学習する
     $L(x) = \text{その階層での再構成} + \text{その階層の } KL(q||p)$ 
  end while
end for

```

---

提案手法のグラフィカルモデルを図 (4.4) に示す. (TODO: 三層以上の場合も図にする), 提案手法は, DSSM の状態変数を N 層に階層化したモデルである. 図 (4.4) は 2 階層の提案モデルを図 (?? TODO) は 3 階層以上の提案モデルを表している. 図 (4.4) の上側の状態変数から一階層の状態変数・二階層の状態変数と呼ぶことになると一階層の状態変数が低次元ベクトル, 二階層の状態変数が高次元ベクトルになっており, 高次元の状態変数の生成・推論時に低次元の状態変数を用いるようなモデルになっている. 高次元の状態変数の遷移時に低階層の状態変数を用いて写像先に関する情報を補助的に与えることで, 学習を安定化させる効果が期待される.

次に提案手法の学習アルゴリズムをアルゴリズム 1 に示す. この学習アルゴリズムは前節「階層的な状態表現の遷移」で述べた, 習熟度に合わせて徐々に高次元の状態ベクトルの遷移を学習するという考え方に基づいており, これにより安定した学習が見込める. 今回簡単のために高階層の潜在表現の学習時にはそれより低階層の状態表現の学習を止めているが, 他の方法も考えられ, これについては考察「低次元状態ベクトルの階層の再学習」で述べる.

### 4.3 類似手法との差分 TODO

本節では提案手法と類似手法の差分について整理する. DSSM を映像生成自体に用いた研究は私の知る限りなく, これは 1 章でも述べたとおり, 自己回帰モデルと比べて高精度な生成には向いていないためだと考えられる. そのため本節では階層性を考慮した既存研究について取り上げる.

- DRAW[] は VAE の潜在変数に階層性をもたせたモデル画像生成で用いられる
- 多層 RNN[] は

- PGAN は

## 第 5 章

# 実験

第 4 章で述べた提案手法の有効性を検証するために, BAIR Push Dataset という行動条件付き映像予測用のデータセットを用いて評価実験を行った. 本章では実験の内容について説明した後に実験結果について述べる.

### 5.1 実験内容

第 4 章で述べた提案手法とベースラインの比較を行う. ベースラインは第二章で述べたシンプルな DSSM とし, 状態ベクトルの次元を 64, 128, 256, 512, 1024 の 6 通りに変えて実験を行う. 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 4096 次元の三階層の状態ベクトルを持つモデルとした. 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. またパラメータの最適化にはそれぞれ確率的勾配降下法アルゴリズム Adam[引用] を用いた. 評価指標には, 定量評価として予測誤差(負の対数尤度)を測り, 合わせて定性評価も行う. またこれらの評価時には, DSSM と同じデコーダー・エンコーダーモデルを持つ VAE を用意し, 時系列方向の遷移を学習する必要がない場合の生成モデルの精度とも比較を行う. (この実験自体 TODO.)

#### 5.1.1 データセット

(TODO: データセットを紹介する図を入れる)

今回用いる BAIR Push Dataset は行動条件付き映像予測と行動条件をつけない映像予測のどちらの研究でも用いられるデータセットであり, カリフォルニア大学バークレー校によって制作・公開されている. [引用] こちらのデータセットは, 様々な物体がおかれた机の上を口

ボットアームがランダムに搔き乱すようにして様々なデータが記録されており、今回はその中から行動系列  $\vec{a}$  と固定視点から観測された画像系列  $\vec{o}$  を用いる。今回用いる行動系列  $\vec{a}$  は、具体的にはロボットのエンドエフェクタの位置姿勢の命令値になっている。観測画像は 64x64 サイズの RGB 画像で、これらのデータは 10hz で撮られている。

### 5.1.2 ベースラインの実装

実装には Facebook 社製の深層学習フレームワークである Pytorch[引用] と、松尾研究室研究員の鈴木さんが中心となって開発されている深層生成モデルライブラリ Pixyz[引用] を主に用いた。学習用データの読み込みとその最適化には Google 社製の深層学習フレームワークである TensorFlow[引用] を用いた。また実装では、DSSM を用いた強化学習手法である PlaNet の公開実装 [引用] を参考にした。まずベースラインの実装を説明した後、提案手法の実装について述べる。

#### モデルアーキテクチャ

ベースラインは 4 つの部分モデルから構成される。

- デコーダー  $p(s_t|x_t)$
- エンコーダー  $q(x_t|h_t)$
- 遷移モデル（事前分布） $p(s_t|s_{t-1}, a_t)$
- 遷移モデル（事後分布） $q(s_t|s_{t-1}, a_t, h_t)$

遷移モデル（事前分布）とデコーダーが生成モデルに相当し、提案手法のグラフィカルモデルの実践部分を表す。遷移モデル（事後分布）とエンコーダーが推論モデルに相当し、提案手法のグラフィカルモデルの点線部分を表している。

#### デコーダー・エンコーダー

(図 TODO) PlaNet に倣い、デコーダー/エンコーダーモデルには WorldModel(Ha 2018) の論文中に記載されているモデルを採用した。デコーダーは各時刻の状態ベクトル  $s_t$  を全結合層で 1024 次元の隠れ変数に変換したあと、4 層の逆畳み込み層で観測画像と同じサイズである 64x64 サイズの RGB 画像に変換する。出力は各ピクセルごとに正規分布をおくが、本研究では簡単のためその分散はそれぞれ 1 に固定している。エンコーダーは各時刻の観測を 4 層の畳み込み層と 1 層の全結合層で 1024 次元の隠れ変数

$h_t$  に変換する。

本研究では、状態変数の次元を変えた際にも隠れ層のパラメータ数などは一切変えなかった。

### 遷移モデル

(図 TODO) 遷移モデルには、ある時刻の状態ベクトル  $s_t$  を生成過程に従って生成する事前分布モデルと、ある時刻の観測  $o_t$  も与えられたときに  $s_t$  を推論する事後分布モデルの 2 種類を用意する。

この 2 つのモデルはアーキテクチャはほとんど共通で入力とするデータだけが違い、事前分布モデルでは一つ前の時刻の状態  $s_{t-1}$  と行動  $a_t$  を入力とするが事後分布ではそれに加えその時刻の観測  $o_t$  をエンコードして得たデータ  $h_t$  を入力とする。出力はどちらのモデルも次の時刻の状態ベクトルの分布の母数(平均と標準偏差)である。アーキテクチャは、入力情報をすべて連結して一度全結合層で変換したものを平均を求める全結合層と標準偏差を求める全結合層のそれぞれで変換し、求めたい分布の平均と標準偏差を出力する。

### 学習の安定化

潜在変数の次元を大きくした際、学習の初期段階でカルバックリライブラリ距離の計算値が発散することがよく起こる。これは遷移モデルが次の時刻の状態ベクトルの分布の標準偏差として 0 に非常に近い値を出力してしまった結果(丸め誤差が発生して)カルバックリライブラリ距離の計算時にゼロ除算が発生してしまうためである。この問題は遷移モデルが出力する標準偏差に下限を設けることで解決することができる。今回の実験では潜在変数の次元を 1024 次元にした際にこのカルバックリライブラリ距離が学習開始後直ちに発散する問題が発生したので、標準偏差の下限値を  $10^{-7}$  とすることで学習をある程度継続することができた。ただし標準偏差に下限を設けることは経験的に学習を難しくすることがわかつっていたので、他の条件での実験の際には適用しなかった。

#### 5.1.3 提案手法の実装

提案手法はベースラインの節で説明した部分モデルをほぼそのまま用いる。二階層目以上の第  $i$  層では、各遷移モデルの入力として一時刻前の状態変数  $s_{t-1}^i$  と行動  $a_t$  だけでなく、一つ低次元の層の同じ時刻の状態ベクトル  $s_t^{i-1}$  も入力とする。また、提案手法では、パラメータ

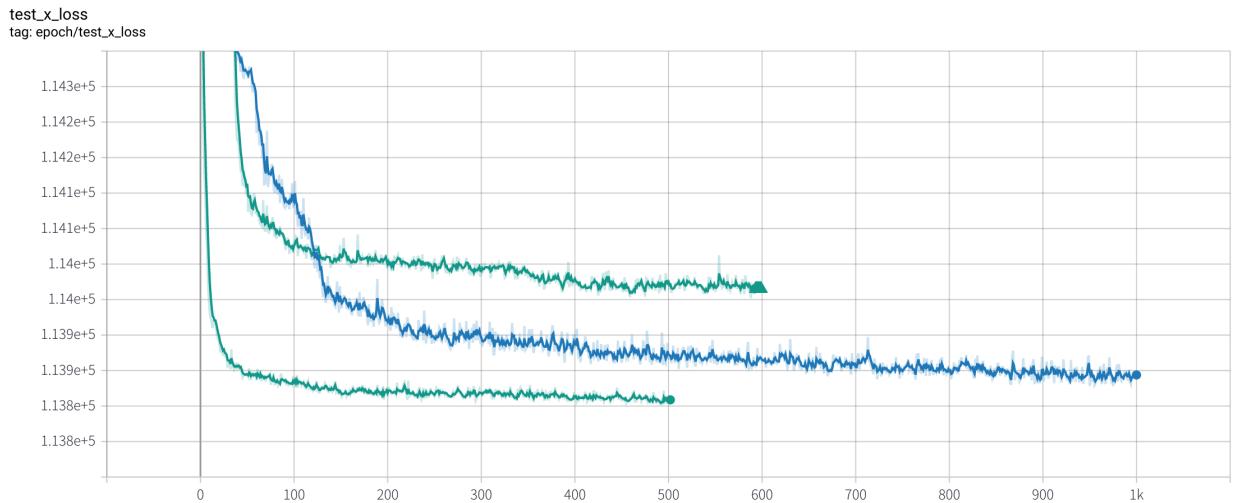


Fig. 5.1 提案手法の学習曲線 一番下が提案手法. TODO: 紹介にする, 凡例入れる

を固定している層の状態ベクトルのサンプリングには、モデルの評価時の設定に合わせて常に事前分布を用いている。これは、パラメータを固定している層はそれ以上学習されないために、事前分布より良い表現が下の階層に渡されることなく、むしろ事前分布で足りない表現を積極的に下の階層の学習で獲得できるようにするためである。その他はベースラインの実装と変えていない。

## 5.2 実験結果

TODO

### 5.2.1 定量評価 (尤度)

上から

- ベースライン (512 次元)
- ベースライン (64 次元)
- 提案手法 (64 + 512 次元)
- (提案手法 (64 + 512 + 4096 次元)(実験待ち))

安定した。

尤度上は改善した

定性的に改善したとはいひ難い結果になった..

高次元で学習できるようになった。これは深層状態空間モデルの大きな問題点を克服できたと言える。

潜在変数のサンプリングが安定し、学習が簡単になったためだと考える。

### 5.2.2 定性評価

# 第 6 章

## 考察

実験結果を受けて、第六章ではより高精度な生成や発展的な問題設定に向けて考察をおこなう。

### 6.1 本研究の貢献

(TODO)

### 6.2 課題

#### 6.2.1 定性的な改善

今回の研究では、尤度上は DSSM の映像予測を改善できたもののあまり定性的には改善が確認できず、特に小さい物体の生成や予測については未だうまく行かなかった。しかし比較用の VAE の実験から、これはそもそもデコーダー・エンコーダーが貧弱で物体一つ一つをはっきりと潜在表現として獲得できていないことが問題であった可能性がある。小さい物体の生成が上手くいかないのはモデルの表現力がそもそも弱かったことが原因であるとして、小さい物体の移動もあまりうまく捉えられていない理由はとして次の 2 つを考えた。

- 生成結果がぼやけたままだと物体の移動などは捉えることが難しい
- そもそも条件付けられる行動が直接的には関与しない物体の移動は、DSSM での学習が難しい

2 が真の理由である場合、より根本的な DSSM の改善を考える必要があるが、素朴には遷移モデルの中間層を増やすことや、行動系列を生成するなどのアプローチが考えられる。また 1

が原因出あった場合には表現力の高いデコーダーを用意するなどするなどで改善が期待できるが、原因は今回の実験からは特定できないためこの問題の解明は今後の課題である。

### 6.2.2 初期状態の推論

今回は初期状態  $s_0$  の推論には、経験的な手法として状態変数の事後分布モデルに初期観測とゼロベクトルを渡することで生成される状態ベクトルを用いたが、これが最適なのかは定かではない。例えば自己回帰モデルを用いる強化学習では、初期状態の推論時に、学習に使う観測系列が始まる直前までの十分な長さの観測系列を使って、ゼロベクトルで表される仮の初期状態を十分な回数更新して（これを burn-in と呼ぶ）真の初期状態として用いることが良いとし、さらに初期状態の推論に用いる観測系列が長いほど精度が向上すると報告した研究がある（R2D2）。本研究ではではロボットの実応用を想定しており、初期状態から予め十分先の将来を予想をした上で行動を開始して欲しいという目的意識があるためこの burn-in は直接適用できないが、同じ初期観測を用いてでも複数回状態ベクトルを更新することは有効である可能性があり、研究の余地がある。

### 6.2.3 低階層の状態表現の再学習

今回提案手法では、複数階層の DSSM モデルを学習させる際には低階層のモデルのパラメータを固定している。しかし学習を簡単にする上ではこれでよいものの、全体としての性能の向上を考えたときにモデルのパラメータを固定することが良いとは必ずしも言えない。今回低階層のモデルは低次元の状態表現で将来の観測の再構成ができるように学習するが、高階層のモデルの学習時には低階層のモデルは再構成しやすい表現を持つ必要はなく、むしろ高階層の状態表現を使った将来予測が簡単になるよう、例えば高階層の状態遷移を補助しやすくなるようなより中心的な状態表現とその遷移を獲得することが求められるはずである。したがって、学習のどこかの段階でモデルの低階層部分を、低階層の状態表現からのデコードをせずに再学習することで、より全体としての性能が高められる可能性が考えられる。

## 6.3 展望

### 6.3.1 敵対的学習の導入

本研究で扱った DSSM では損失関数に二乗誤差を利用しているため、画像空間上の小さな特徴量、例えば小さい物体や背景に近い色をした物体などは無視されやすい。二章で取り上げ

た VAE や DSSM など最尤推定を用いる深層生成モデルは目的関数に生成誤差を取るために一般的に共通してこのような問題を持つ。この問題は、近年高品質な深層生成モデルとして研究が盛んな敵対的生成ネットワーク、GAN(generative adversarial networks) を補助的に用いることによる解決などが考えられる。

### 6.3.2 多視点からの映像予測

本研究では 1 視点からの映像予測を行っているが、ロボット実機に映像予測を応用する際にはロボットの視点が動くような問題設定や、より高性能な予測を行うために多視点からの観測が使える前提で映像予測を行うというような問題設定なども考えられる。深層学習研究で複数視点からの観測を扱う問題の先行研究として Generation Query Network(GQN)<sup>[1]</sup> があり、複数視点からの観測とその視点位置が与えられたときに別の視点からの観測を予測して生成する問題を解いている。この研究では視点不变な共通の潜在変数の獲得を行い、クエリとして視点情報が渡されたときにその視点での観測を返すモデルを考えている。DSSM に GQN のアプローチを組み合わせることは、より良い表現の獲得と実用性の向上につながるため重要であり、実際 Temporal GQN などこの方向性の研究はすでに進められているが、このような問題設定においても本研究の提案手法は有効であると考えられるので検証を行いたい。

### 6.3.3 他の構造を持つ潜在表現との併用

今回の提案手法では潜在変数の階層性のみを仮定してモデルを構築したが、近年の映像予測モデルのほとんどは潜在変数に平面的な構造を仮定し、状態表現をベクトルではなく行列として扱うことで性能の向上を実現している [high fide, SVG]。状態表現に階層性をもたせるこの有効性は実験で示したとおりであるが、他にも環境内でより立体的な行動を考える際には潜在変数に立体性を仮定したり、さらに例えば操作対象の物体が 1 つで構造が既知の場合には操作対象のメッシュ情報やグラフ構造を状態表現の構造として用いることも可能であると考えられる。このようなベクトル以外の構造を持つ状態表現を学習したい場合においても、低次元のベクトル状の潜在表現を用いて階層的なモデルを考えることで精度を高めたり学習をスムーズに進められるようになる可能性があり、本研究の提案アルゴリズムが活かせると考える。

### 6.3.4 メタ学習の導入

機械学習研究において、似たタスク集合が与えられたときにそれらのタスク全てに汎化できるようなモデルの獲得を目指すメタ学習という研究分野がある。行動条件付き映像予測では与えられたデータから環境の遷移を学習するが、遷移自体が同じで観測が異なるような場合、例えば本研究で扱った BAIR Push Dataset をベースに考えると机や操作対象の物体、そしてロボットの操作方策は同じままロボットの機種だけが異なるような場合にも適切に映像予測がしたいというような問題設定が考えられるが、このような設定ではメタ学習手法を用いることで様々なロボットに対応できる映像予測が可能になる。この場合具体的には、状態表現にロボットの機種情報が含まれないよう状態表現とロボットの機種クラスの相互情報量を最小化する制約を置いて遷移を学習することでロボットの機種に依存しない遷移モデルを獲得することができ、代わりにデコーダーにロボットの機種クラス情報を追加で渡すことによって適切に映像予測が可能になると予想できる。

### 6.3.5 映像予測用データ収集の人による代替

映像予測を行うにはまずデータを収集する必要があるが、映像予測の実ロボットでの応用を考えた場合、BAIR Push Dataset のようにランダムにロボットを動かしてデータを集めただけでは本来不十分である。理想的には、実ロボットが環境中である程度期待する動きをすることがわかっている場合にも、期待する行動系列をまんべんなく試行してデータを集めたい。しかしそもそもロボットが環境を適切に操作するための制御方策を獲得すること自体難しいことが多い、適切に映像予測用のデータが集められるような良い制御方策が獲得できていればわざわざ映像を予測する必要はないかもしれない。そこで、考えられるのがデータの収集を人間が行う方法で、人間であれば環境内で期待する様々な行動系列実行することは容易い。ここで以下の二つの問題が発生するがこれらは近年の研究や技術で簡単に解決することが見込める。

- ロボットとは異なり、人間の行動系列情報を取得することが難しい
- データセットでは人間が、実環境ではロボットが操作を行うため、適切に映像予測を学習できても実機ロボットの映像予測に利用できない

まず一点目については、これは操作主体の人をカメラや工学センサー、磁気センサーなどを使ったモーショントラッキングやモーションキャプチャの技術によって人の行動系列（この場

合人の運動)を計測することができる。二点目については、例えばTCNで提案されているような手法で、同じ操作を行う人間の行動系列データとロボットの行動系列データの対応関係を予め学習した上で、データセットとロボットの実観測から人やロボットが写り込んでいる部分を削除してこれまで通り遷移モデルを学習・利用するアプローチが考えられる。

このようにしてロボットの代わりに人がデータを集めることで、非剛体物の操作などより複雑な映像予測の問題設定を扱うことが可能になると見える。

## 6.4 社会応用

本研究ではDSSMの拡張を考えたが、DSSMあるいは映像予測の研究の社会応用の可能性を述べる。

### 6.4.1 実機ロボットへの応用

(考え中 TODO)

### 6.4.2 物理シミュレーションの近似

本研究で扱った深層学習ベースの行動条件付き映像予測では物理現象の結果として観測されるデータの予測を行うが、これは物理現象の近似という意味で現行の物理シミュレーションソフトウェアと共に目的を持っていると解釈ができる。現在、物理シミュレーションの手法としては、既知の様々な物理法則を記述し微小時間・微小空間単位で逐次的に各領域の状態を計算して全体の結果を予測することが一般的である。しかしこのアプローチは、物理法則が既知である必要があり、また複雑な物理法則に対しては予測に膨大な時間がかかりリアルタイムに予測ができないという問題がある。例えば[文献]は、蕎麦にオイスターソースをかけて混ぜる物理シミュレーションを扱っているが、1フレームごとの見た目はとてもリアルなもののが現象としては依然不自然な部分があり、さらに30fpsで1秒の予測をするのに29時間かかると報告している。現行の物理演算ベースの映像予測に対し深層学習ベースの映像予測は、物理法則の正しさや多視点から見た際の一貫性が保証できないなど機能として制限は多いものの、必ずしも物理現象が既知である必要はなく、また一度学習すればリアルタイムで予測を行うことができる。さらに、必要なデータを集めることで機能の制限を解消することもできるはずである。このようなことから、映像予測手法は複雑な物体の物理シミュレーションの機能を部分的に代替できる可能性があり、これによってシミュレートできる物体や現象の幅が広げられる

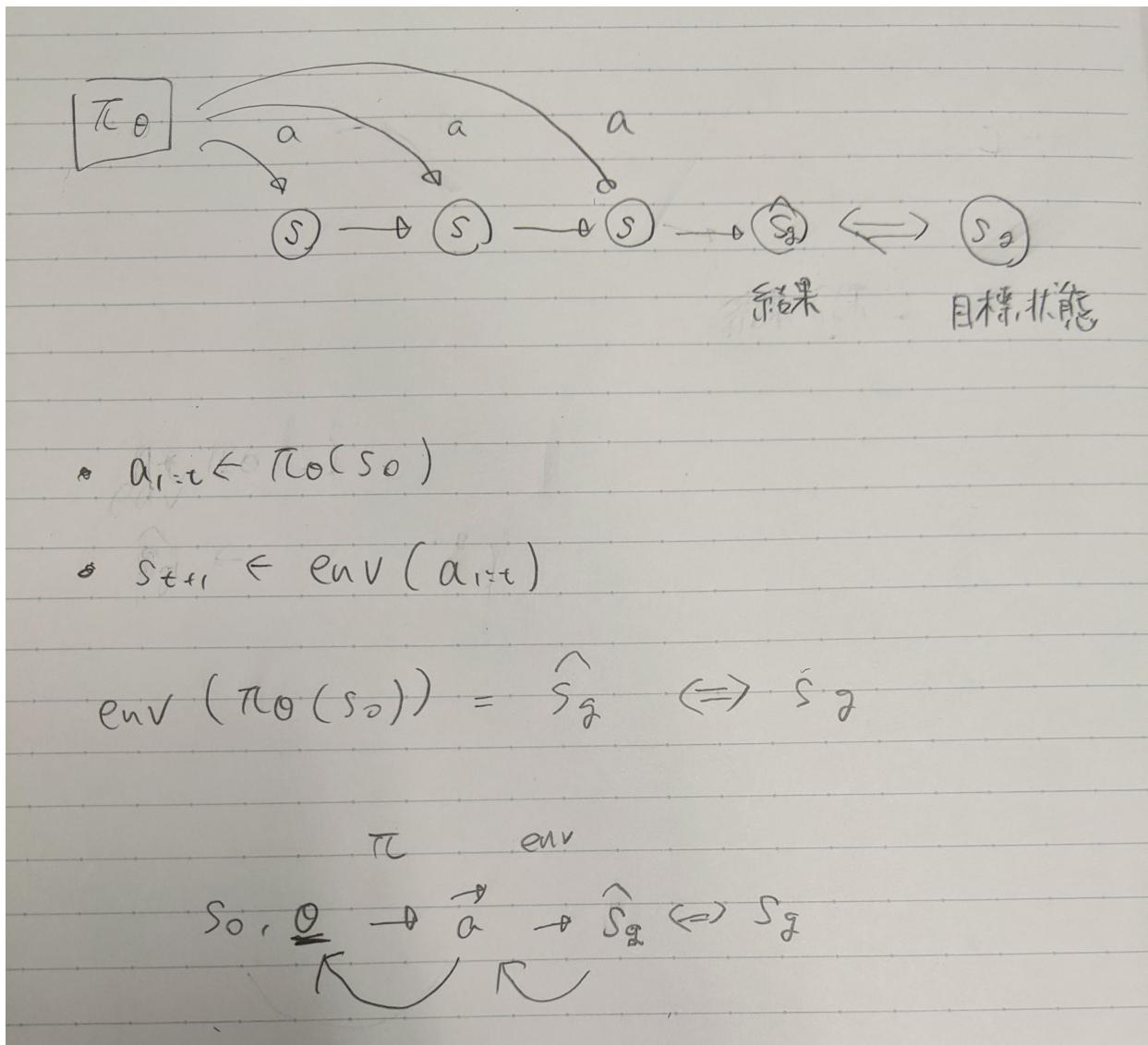


Fig. 6.1 TODO:きれいにする

と考える。

#### 6.4.3 微分可能な環境モデルとしての利用

深層学習を用いてロボットの制御方策を学習するロボット学習において、基本的にロボットは実環境かシミュレーション環境で行動を起こすまでは結果を観測することができない。これは、ロボットの行動を入力、その結果の観測を出力と考えると、入力と出力の間に環境という未知の関数が存在し、得られた出力から入力を最適化する際に直接勾配法を用いることができないことを意味する。深層強化学習ではこの問題を様々なアプローチで解決しているが、それ

でも環境中で多くの試行を重ねる必要があり学習に時間がかかる上、エージェントの経験の偏りによって獲得される方策が変わってくるため性能を安定して上げることは難しい。近年の深層強化学習研究では環境との相互作用からエージェントが自ら環境のモデル自体を明示的に学習し、この学習された環境（世界モデル／内部モデル）を代わりに使うことで実環境での試行回数を減らし学習効率を高めるアプローチが精力的に研究されている。この世界モデルと同じような発想で、実機ロボットにおいて行動と結果の間に環境という未知の関数があるならば、環境という関数自体を微分可能なモデルで近似できれば、ロボットの行動方策を直接勾配法で最適化できるはずであるとする考えがある [引用]。 (TODO 図を入れる) 具体的には、近似した環境モデル  $env$  と最適化したい制御方策  $\pi$  を用いてロボットがある初期状態  $s_{start}$  からある目標状態  $s_{goal}$  までの行動をプランニングしたときに、環境モデル  $env$  によって予測される結果状態  $\hat{s}_{goal}$  との誤差を小さくするように勾配法で直接方策を学習する。

$$env(\pi(s_{start})) = \hat{s}_{goal} \Leftrightarrow s_{goal}$$

ここで、本研究のようなディープラーニングベースの環境の近似モデルは微分可能であるため、この式で示す方法での方策の最適化に用いることができる。特に環境が既知で、予め他のエキスパートロボットやあるいは人間によって多くの試行を重ねてデータの収集ができる、環境を非常に高い精度で近似できる場合には、この近似モデルを使った方策の学習は非常に有効であると考えられる。このように映像予測手法が発展することによって、全く新しいアプローチのロボット学習の研究が可能になり、近似環境モデルを用いた素早く安定した学習によってロボットの応用の可能性が広がると考えている。

## 第7章

### 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。

## 7.1 TODO: 謝辞

## 7.2 その他書いたほうがいいこと

- ただの VAE との比較 (Enc Dec は同じ条件で)
- 30 フレーム先の予測
- FVD スコア?

# 謝辞

本論文を作成するにあたり、多くの方々にご協力をいただきました。

最後に、配属からの一年間、あらゆる面でサポートをしてくださった松尾研究室の皆様に御礼申しあげて、謝辞とさせて頂きます。

東京大学工学部システム創成学科

知能社会システムコース

松尾研究室学部 4 年

平成 31 年 2 月 近藤生也

## 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*, pp. 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] Emily Denton and Rob Fergus. Stochastic video generation with a learned prior, 2018.
- [4] Ruben Villegas, Arkanath Pathak, Harini Kannan, Dumitru Erhan, Quoc V. Le, and Honglak Lee. High fidelity video prediction with large stochastic recurrent neural networks, 2019.
- [5] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013.