

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

概要

近年、深層ニューラルネットワーク (DNN) を使用した機械学習手法の発展を背景とし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いる。ロボット学習において未来の観測、特に自らの行動に対するフィードバックを適切に予測するよう学習することは、環境の遷移モデル (内部モデル) を獲得すること、そしてそれを用いたプランニングや行動結果の予測、安全性の評価などに欠かせない。時間変化する環境の DNN を用いたモデルとして深層状態空間モデル (Deep State Space Model) があり、これは昨今深層強化学習の分野で取り入れられているが、深層状態空間モデルを用いて映像予測を学習することで内部モデルが獲得できるとされている [1]。しかし現在一般的に使われて深層状態空間モデルとその学習方法は各時刻の状態が低次元のベクトルで表現できるという仮定を暗に置いており、高次元な表現を仮定すると学習が進まず、複雑な環境の予測問題にスケールさせる方法は自明でない。

本研究ではまずこのような問題があることを実験的に指摘した上で、状態の表現に階層性を考えることで深層状態空間モデルにより高次元な状態表現を仮定できるように拡張し、低次元な表現を仮定した場合よりも良い表現が獲得できることを示す。評価実験では様々な物体が置かれた机上でのマニピュレーションを題材にしたデータセットを用いてロボットの行動で条件付けた映像予測を行う。単純な深層状態空間モデルをベースラインに設定し、提案手法の定性的・定量的な有効性を示す。

目次

概要	i
第 1 章 序論	1
1.1 本研究の背景	1
1.1.1 ロボット学習	1
1.1.2 状態表現学習	1
1.2 本研究の目的	2
1.3 本論文の構成	2
第 2 章 前提知識	3
2.1 変分自己符号化器 (VAE)	3
2.2 深層状態空間モデル (DSSM)	6
第 3 章 深層状態空間モデルの限界	9
3.1 学習が失敗した例	9
3.2 深層状態空間モデルの問題点	9
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	11
4.1 問題設定	11
4.2 ベースライン	11
4.3 提案手法	11
4.3.1 確率モデル・最適化	11
4.3.2 モデルアーキテクチャ	11
4.4 類似手法との差分	11
第 5 章 実験	12

5.1	実験内容	12
5.1.1	実験概要	12
5.1.2	BAIR Push Dataset	12
5.1.3	hogehoge スコア	12
5.1.4	実験条件	12
5.1.5	学習時のテクニック	12
5.2	実験結果	12
5.2.1	定量評価 (尤度)	12
5.2.2	定量評価 (hogehoge スコア)	13
5.2.3	定性評価	13
第 6 章	考察	14
6.1	本研究の貢献	14
6.2	今後の課題	14
6.2.1	転移学習	14
6.3	社会応用	14
6.3.1	実機ロボット	14
6.3.2	物理シミュレーションの近似	14
第 7 章	まとめ	16
謝辞		17
参考文献		18

目次

2.1	VAE のグラフィカルモデル	3
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.	4
2.3	VAE を用いて生成された画像の例 ([2] より引用)	5
2.4	SSM のグラフィカルモデル	7
2.5	DSSM で生成される映像の例. ただし SSM を少し拡張した RSSM モデルを 使っている	7
2.6	DSSM で生成される映像の例.	8
3.1	DSSM の学習が失敗した例	10

表目次

第 1 章

序論

1.1 本研究の背景

1.1.1 ロボット学習

多様な環境でさまざま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、少しずつ遂行可能なタスクが増えている。ロボット学習は問題設定によって様々な手法があるが、例えば強化学習をベースにして複雑な形状の物体の把持方策を (grasp2vec) 学習するものや、模倣学習によって人のデモンストレーションを真似て食器の出し入れを学習するもの (TCN)、更には DNN ベースを含め様々な手法を統合して家庭用ロボットによる片付け [3] を遂行する手法などが提案されている。このように実用的なタスクも解決できつつあり、自動化が進む社会においてその活躍の期待値が高まっている。

1.1.2 状態表現学習

(<https://arxiv.org/pdf/1802.04181.pdf>) データを特徴づける情報を抽出する DNN は特徴抽出が得意。教師なし学習で表現学習と呼ばれる。時間変化する環境に拡張した表現学習のことで、各時刻の生の観測データやその系列からこの状態表現 (特徴) を見つけることが目的となる。

planet dreamer 上記のような強化学習や実機タスクへの応用から (行動条件付き) 映像予測だけを切り取った研究も多い。

1.2 本研究の目的

提案手法の有効性について定性的・定量的な評価を行う。

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とおりである。

では

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 $\mathbf{z} \in \mathbb{R}^m$ があるとし，Fig. 2.1 のようなグラフィカルモデルで表現される確率モデルを考えた上で式 (2.2) で表されるデータ \mathbf{x} の確率的生成過程をニューラルネットワークによってモデル化する．

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

Fig. 2.1 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説に基づいている．また，潜在表現 \mathbf{z} を考えると，正しい $P(\mathbf{x}|\mathbf{z})$ が得られれば， \mathbf{z} が与えられたときに $P(\mathbf{x}|\mathbf{z})$ を使って



Fig. 2.1 VAE のグラフィカルモデル

それに対応する \mathbf{x} を生成することができ、 $P(\mathbf{z}|\mathbf{x})$ を考えると \mathbf{x} が与えられた際の良い表現としての \mathbf{z} が得られる。

VAE では、以下の 2 つの仮定を置き、さらに $p(\mathbf{z}|\mathbf{x})$ を近似する $q(\mathbf{z}|\mathbf{x})$ を導入する。 $p(\mathbf{x}|\mathbf{z})$, $q(\mathbf{z}|\mathbf{x})$ 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データ \mathbf{x} 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。ただし $p(\mathbf{z}|\mathbf{x})$ を直接考えないのは、 $p(\mathbf{x}|\mathbf{z})$ を表すニューラルネットのパラメータで $p(\mathbf{z}|\mathbf{x})$ を表すことが難しいためである..

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

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

式 (2.2) は、潜在空間が標準正規分布に従っているという仮定であり、式 (2.3) は、 \mathbf{x} に条件づけられた潜在変数の分布が正規分布に従うという仮定である。 $q(\mathbf{z}|\mathbf{x})$ は推論分布/推論モデルと呼ばれ、グラフィカルモデル中に記述すると Fig. 2.2 となる。

式 (2.2) の対数尤度をとることで、以下の変分下限を導出する、

$$\begin{aligned} \log p(\mathbf{x}) &= \log \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z} \\ &= \log \int q(\mathbf{z}|\mathbf{x}) \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{q(\mathbf{z}|\mathbf{x})} d\mathbf{z} \\ &\geq \int q(\mathbf{z}|\mathbf{x}) \log \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{q(\mathbf{z}|\mathbf{x})} d\mathbf{z} \end{aligned} \quad (2.4)$$

$$\begin{aligned} &= \int q(\mathbf{z}|\mathbf{x}) \log p(\mathbf{x}|\mathbf{z}) d\mathbf{z} - \int q(\mathbf{z}|\mathbf{x}) \log \frac{q(\mathbf{z}|\mathbf{x})}{p(\mathbf{z})} d\mathbf{z} \\ &= \mathbb{E}_{\mathbf{z} \sim q(\mathbf{z}|\mathbf{x})} [\log p(\mathbf{x}|\mathbf{z})] - D_{\text{KL}}(q(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})) \end{aligned} \quad (2.5)$$

ここにはイェンセンの不等式を用いる。また、数値計算をする都合上式の \mathbf{z} の周辺化は難

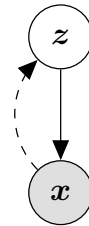


Fig. 2.2 推論分布を導入した VAE のグラフィカルモデル。実線は生成分布、点線は推論分布を表す。

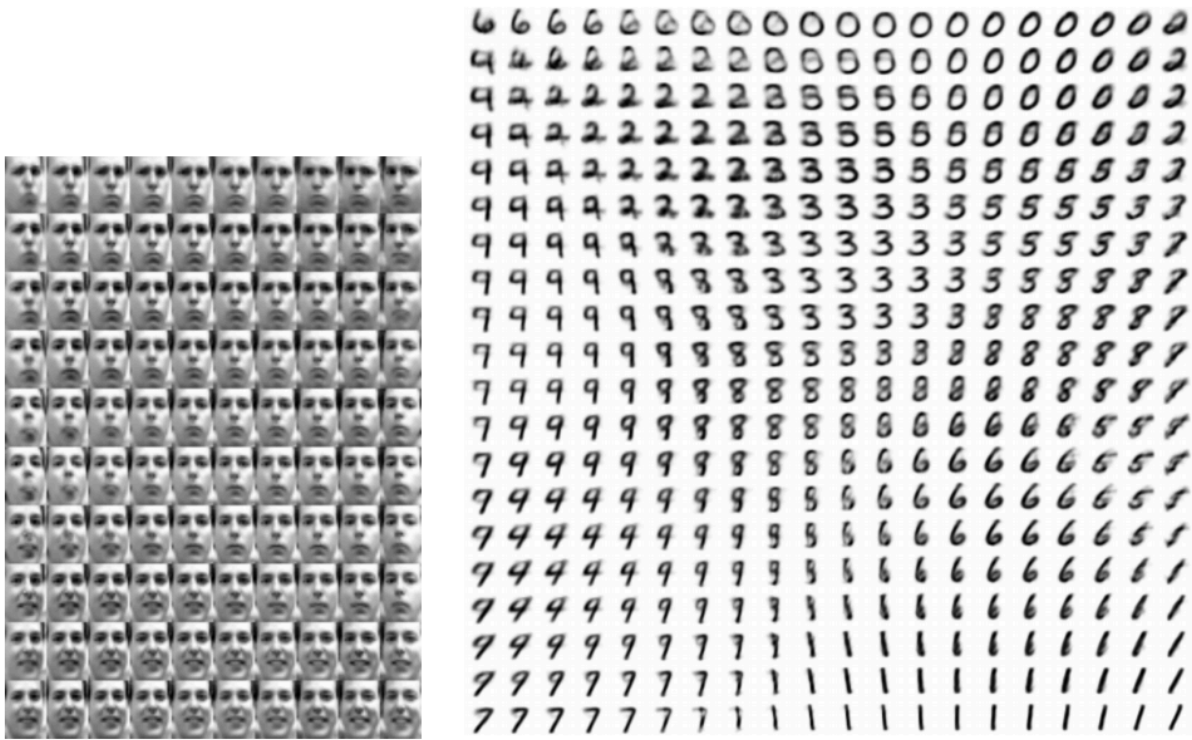


Fig. 2.3 VAE を用いて生成された画像の例 ([2] より引用)

しいため、天下りの的に $q(z|x)$ を導入する．式では、 z の周辺化を解析的には計算できないため、 $q(z|x; \phi)$ 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．ただし通常 $L=1$ で計算される．またこの近似は、式で $p(z)$ から z をサンプリングすることと比較して現実的な計算コストに抑えられるに注意されたい．式 (2.5) の第 2 項の D_{KL} (カルバックライブラー距離) は、いま $p(z)$, $q(z|x; \phi)$ 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る．式 (2.5) 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して NN を学習させる．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. 2.3 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 によって生成された画像、右は MNIST で学習した VAE の生成画像である．

2.2 深層状態空間モデル (DSSM)

深層状態空間モデル (DSSM, Deep Karman Filters, Deep Markov Model などとも呼ばれる) は、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状態表現が遷移すると考えたモデルであり、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。

深層強化学習の分野では、エージェントが環境に行動を起こした結果に観測が得られるようなときに、行動と環境の変化をモデル化する際に用いられる。本論文では一般的な以下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置いても以下の説明は同様にして考えることができる。DSSM では VAE と同じように、以下の変分下限から最尤推定によって $p(o|a)$ や $p(s|x)$ を得る。

$$\begin{aligned}
\log p(o_{1:T}|a_{1:T}) &= \log \prod_{t=1}^T p(o_t|a_{1:t}) \\
&= \log \prod_{t=1}^T \int p(o_t|s_t)p(s_t|s_{t-1}, a_t) ds_t \\
&= \sum_{t=1}^T \log \int p(o_t|s_t)p(s_t|s_{t-1}, a_t) ds_t \\
&= \sum_{t=1}^T \log \int q(s_t|s_{t-1}, a_t, o_t) \frac{p(o_t|s_t)p(s_t|s_{t-1}, a_t)}{q(s_t|s_{t-1}, a_t, o_t)} ds_t \\
&\geq \sum_{t=1}^T \int q(s_t|s_{t-1}, a_t, o_t) \log \frac{p(o_t|s_t)p(s_t|s_{t-1}, a_t)}{q(s_t|s_{t-1}, a_t, o_t)} ds_t \\
&= \sum_{t=1}^T \left(\int q(s_t|s_{t-1}, a_t, o_t) \log p(o_t|s_t) ds_t \right. \\
&\quad \left. - \int q(s_t|s_{t-1}, a_t, o_t) \log \frac{p(s_t|s_{t-1}, a_t)}{q(s_t|s_{t-1}, a_t, o_t)} ds_t \right) \\
&= \sum_{t=1}^T \left(\mathbb{E}_{s_t \sim q(s_t|s_{t-1}, a_t, o_t)} [\log p(o_t|s_t)] \right. \\
&\quad \left. - \mathbb{E}_{s_{t-1} \sim q(s_{t-1}|s_{t-2}, a_{t-1}, o_{t-1})} [\text{D}_{\text{KL}}(q(s_t|s_{t-1}, a_t, o_t) \| p(s_t|s_{t-1}, a_t, o_t))] \right)
\end{aligned} \tag{2.6}$$

ただし s_0 は、初期状態が一定な問題を考える際は常にゼロベクトルなど一定の値にするこ

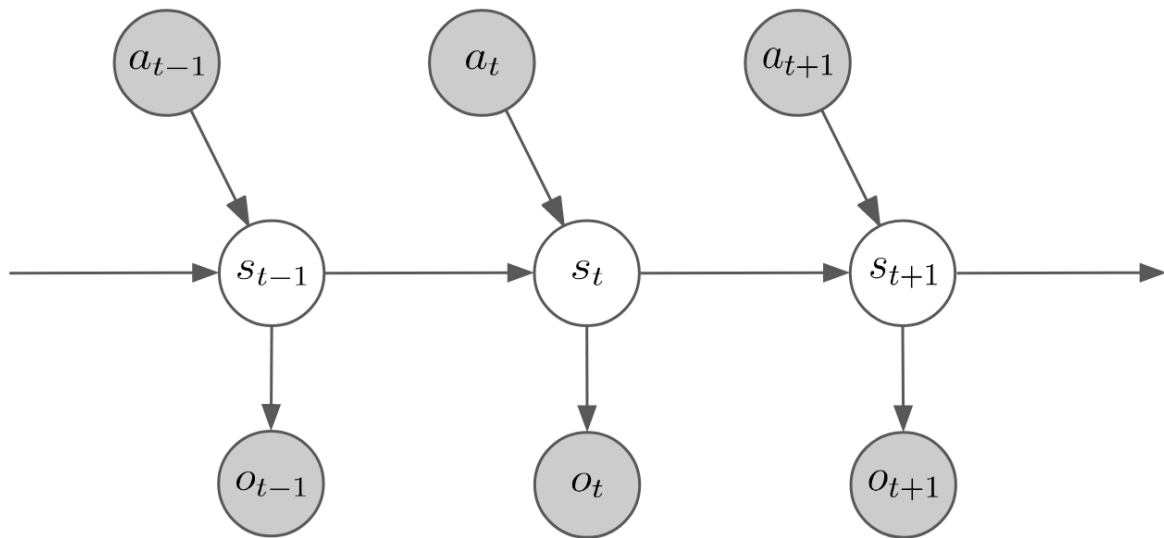


Fig. 2.4 SSM のグラフィカルモデル

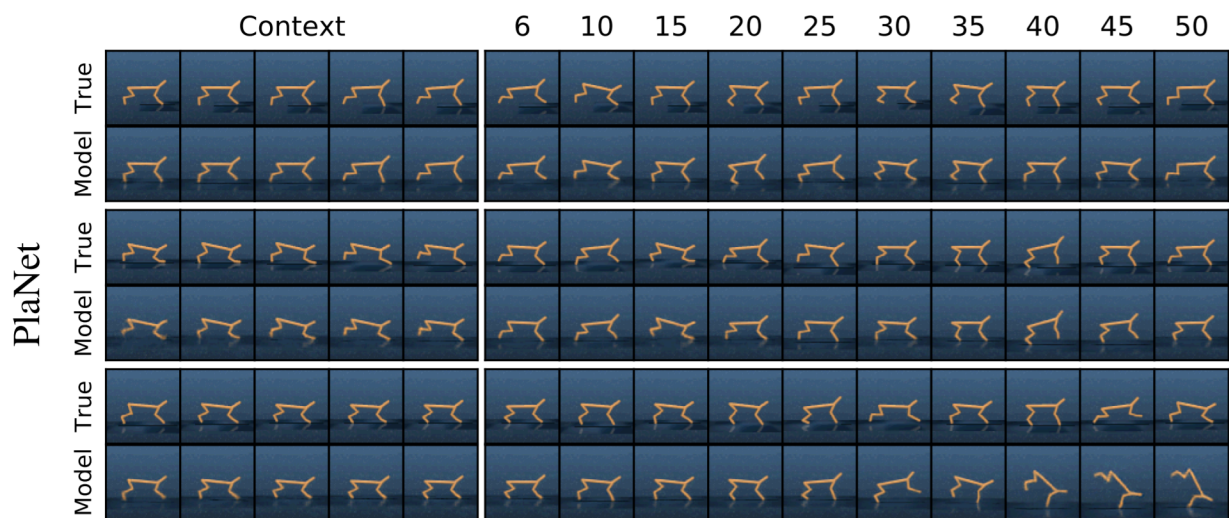


Fig. 2.5 DSSM で生成される映像の例. ただし SSM を少し拡張した RSSM モデルを使っている

ともあるが、そうでない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、(2.6) の第 2 項の s_0 のサンプリングはこれで置き換える．

式は、イェンセンの不等式を使い、 q は p の近似である．

以上が DSSM の概要である．DSSM を用いると、2.5, 2.6 のような新たなデータを生成することができる．action として回転情報が入れている．

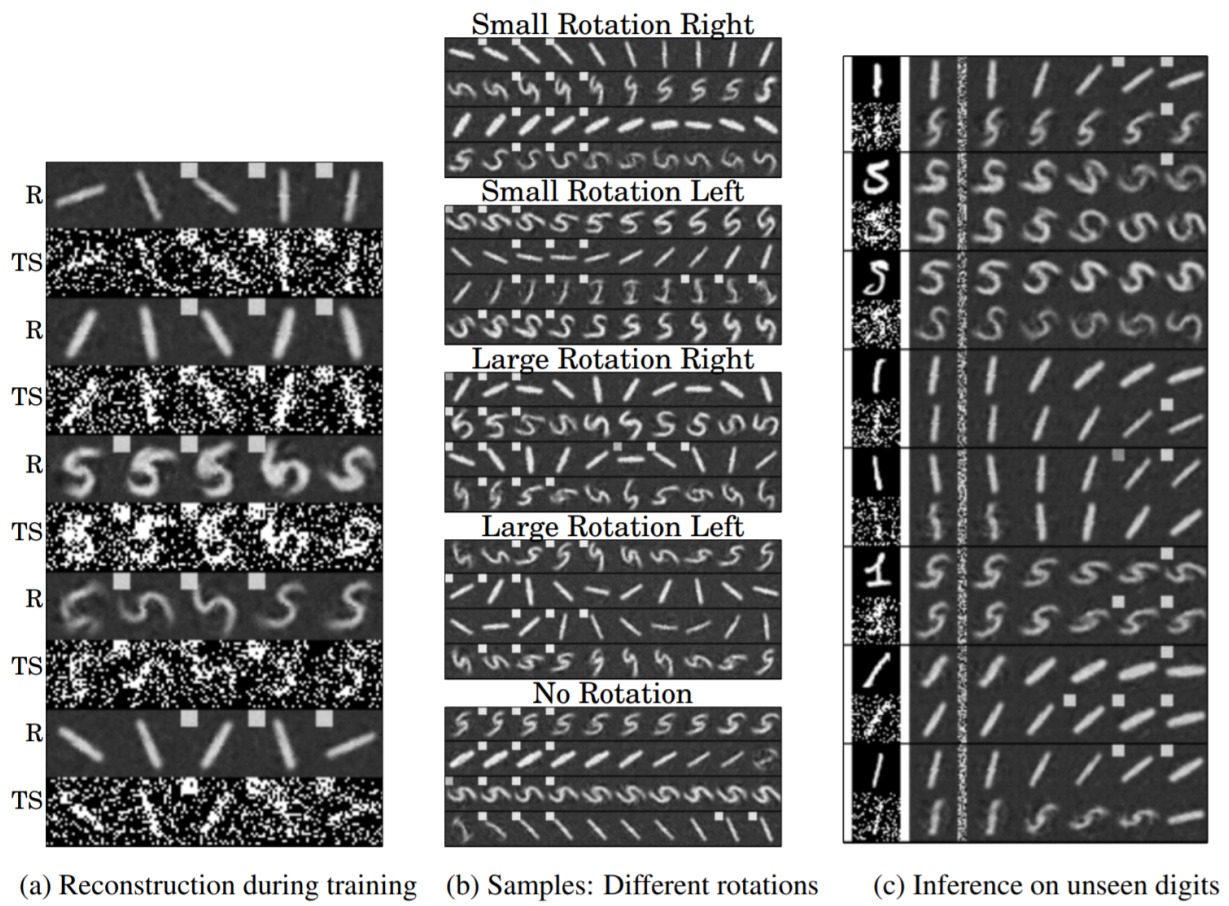


Fig. 2.6 DSSM で生成される映像の例.

第 3 章

深層状態空間モデルの限界

深層状態空間モデルは、潜在変数の次元を大きくしたときの学習が難しい．そのことを実験的に示した上で、理論的な問題点を述べる．

3.1 学習が失敗した例

3.1 は、本論文の 4 章以降でベースラインとして述べる通常の SSM モデルを、潜在変数の次元を何通りかに変えてモデルを構築し、学習中の様子ある．潜在変数を低次元にすると順調に学習が進み、少しずつ近い映像が出力されるようになるが、次元を大きくすると途中で目的関数とする変分下限の値が正の方向に発散しやすくなる．さらに調べると、変分下限のうち

3.2 深層状態空間モデルの問題点



Fig. 3.1 DSSM の学習が失敗した例

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

を目指す。

はじめに、提案手法で扱う問題設定について整理し、その後提案手法の具体的な説明として、確率モデル・最適化とモデルアーキテクチャについて述べていく。

4.1 問題設定

4.2 ベースライン

4.3 提案手法

4.3.1 確率モデル・最適化

4.3.2 モデルアーキテクチャ

4.4 類似手法との差分

DRAW[] は遷移がない，多層 RNN[] は直前の観測を逐次的に入力する．[] で RNN ベースはダイナミクスの獲得が弱く，長期の予測が難しいとされている．FHMM は離散低次元で EM アルゴリズムによって最適化をしている (video Flow...)

第 5 章

実験

本研究では，第 4 章で述べた提案手法の有効性を検証するために， で使用されている Shepard Metzler データセットを用いて，評価実験を行った．本章では，行なった実験の内容について説明した後，実験結果について述べていく．最後に実験結果を踏まえた考察を行う．

5.1 実験内容

5.1.1 実験概要

adam 評価指標には，予測誤差 (負の対数尤度)，hogehoge スコアを用いる push dataset を使う

5.1.2 BAIR Push Dataset

5.1.3 hogehoge スコア

オリジナル/ノンオリジナル

5.1.4 実験条件

5.1.5 学習時のテクニック

5.2 実験結果

5.2.1 定量評価 (尤度)

安定した．

潜在変数のサンプリングが安定し、KL 項が安定して下がりやすくなったからだと思う
尤度が下がった

5.2.2 定量評価 (hoge hoge スコア)

5.2.3 定性評価

きれいになってくれ

第 6 章

考察

6.1 本研究の貢献

6.2 今後の課題

6.2.1 転移学習

環境のダイナミクスとして、普遍的な物理法則を一部獲得していると考えられることから、様々なタスクで特に階層的なモデルでは、低次元の state に大域的な物理法則に関するダイナミクス、高次元の state に比較的表面的なダイナミクスを獲得していると考えられ、普遍的な環境の多少の変化には

6.3 社会応用

6.3.1 実機ロボット

非剛体物体のマニピュレーションが考えられる。

6.3.2 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近似を行うが、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同じことを行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、既知の様々なスケールの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 [1] 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物理現象としては以前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。

これからの人工知能の研究には仮想現実環境の開発が欠かせない。このことは、昨今強化学習やロボット学習の分野で相次いで世界中の研究機関が学習用のシミュレーション環境 [Meta world, control suit, RLBench] を開発し公開していることから伺える。物理的なタスクを解けるよう学習するには実環境では危険であったり学習の並列化が難しいため、先にシミュレーション内で学習することが有効である [sim2real の論文]。また、与えられたタスクやタスクの集合を解くようトップダウン型の人工知能研究とは対比的に、動物の知能の仕組みや知性の獲得の過程を再現するボトムアップ型の人工知能研究も大切であり [考える脳]、ボトムアップ型人工知能研究にも仮想現実環境が必要であると考えられる (The Two Faces of Tomorrow, SAO)

これらの人工知能研究の動向・方向性を踏まえ、物理演算ベースと深層学習ベースの物理シミュレーションの融合、あるいは深層学習ベースによる代替を模索することも重要になるだろう。

第 7 章

まとめ

本論文では、はじめに について取り上げ、その問題点として である点を挙げた。そして、その問題点の原因として、 を指摘し、 を行なった。

そこで、本論文では、これらの問題点を解消する改善手法の提案を行なった。提案手法では、 することで、 し、 を可能にした。

実験では、BAIR Push Dataset を用いてベースライン手法を比較し、定性的・定量的な評価を行なった。提案手法は、 においてベースライン手法である を上回る結果となり、提案手法が が実験的にも示された。

最後に考察として、今後の課題や社会応用について議論し、本研究が深層生成モデルを用いて観測情報から環境をモデル化する世界モデルの研究の 1 つとして位置付けられることを述べ、今後の研究の方向性についての展望を整理した。

本研究は、既存の工学的な技術の代替としてのみならず、より汎用的な人工知能技術の達成のために重要な技術の 1 つであると考えており、今後も実環境への提案手法の適用や、時系列遷移を考慮したモデルの考案、強化学習技術との融合などを見据えて、人工知能分野の研究の発展に貢献していきたいと考えている。

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，謝辞とさせていただきます．

東京大学工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] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623*, 2015.
- [2] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. In *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2013.
- [3] Jun Hatori, Yuta Kikuchi, Sosuke Kobayashi, Kuniyuki Takahashi, Yuta Tsuboi, Yuya Unno, Wilson Ko, and Jethro Tan. Interactively picking real-world objects with unconstrained spoken language instructions. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3774–3781. IEEE, 2018.