

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
1.1.3 映像予測	1
1.2 本論文の構成	1
第 2 章 前提知識	2
2.1 変分自己符号化器 (VAE)	2
2.2 深層状態空間モデル (DSSM)	5
第 3 章 深層状態空間モデルの限界	6
3.1 学習が失敗した例	6
3.2 深層状態空間モデルの問題点	6
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	7
4.1 問題設定	7
4.2 ベースライン	7
4.3 提案手法	7
4.3.1 確率モデル・最適化	7
4.3.2 モデルアーキテクチャ	7
4.4 類似手法との差分	7
第 5 章 実験	8

5.1	実験概要	8
5.2	データセット	8
5.3	実験条件	8
5.4	評価指標	8
5.4.1	hogehoge スコア	8
5.5	学習時のテクニック	9
5.6	実験結果	9
5.6.1	定量評価 (尤度)	9
5.6.2	定量評価 (hogehoge スコア)	9
5.6.3	定性評価	9
第 6 章	考察	10
6.1	本研究の貢献	10
6.2	今後の課題	10
6.2.1	転移学習	10
6.3	社会応用	10
6.3.1	ロボットマニピュレーション	10
6.3.2	物理シミュレーションの近似	10
第 7 章	まとめ	11
謝辞		12
参考文献		13

目次

2.1	VAE のグラフィカルモデル	2
2.2	VAE を用いて生成された画像の例 ([2] より引用)	4

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を実行することのできる，汎用的なロボット（generalist robots）（Escudero Rodrigo et al., 2015; Finn, Yu, et al., 2017）の開発は，ロボット工学の最重要課題の一つである．

1.1.2 状態表現学習

DNN は特徴抽出が得意．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．

1.1.3 映像予測

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．

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．

1.2 本論文の構成

本論文の構成は以下のとおりである．

第 2 章

前提知識

本章では，本研究の前提となる知識に関して説明する．具体的には，代表的な深層生成モデルである変分自己符号化器 (Variational Auto-encoder, VAE) とその拡張である DRAW についてまとめた後，本研究のベースラインとなる生成クエリネットワークについて説明する．

最後に，生成クエリネットワークの確率モデルの考察において用いるメタ学習と呼ばれる機械学習のフレームワークについて説明する．

2.1 変分自己符号化器 (VAE)

変分自己符号化器 (Variational Auto-encoder, 以下 **VAE**)[2] は，深層生成モデルの一種である．VAE では，データ $\mathbf{x} \in \mathbb{R}^n$ はある潜在変数 $\mathbf{z} \in \mathbb{R}^m$ から生成されると考え，その確率的生成過程 $p(\mathbf{x}|\mathbf{z})$ を多層ニューラルネットワークを用いてモデル化する．つまり，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仮定し，そのパラメータ θ をニューラルネットワークのパラメータとして表現する．データ \mathbf{x} は手書き数字の画像に該当するが，これはそれぞれの要素が 1 つのピクセルの値に相当する $28 \times 28 = 784$ 次元のベクトルで表された高次元な表現である．潜在変数 \mathbf{z} は，データ \mathbf{x} をより低次元に表現する．これは，画像

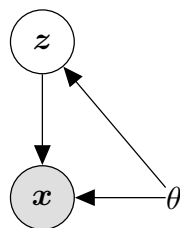


Fig. 2.1 VAE のグラフィカルモデル

のような高次元なデータは画像空間上の非常に限られた領域に局所的に存在しており、それらはより低次元に表現可能であるとする多様体仮説に基づいている。

すると、データの分布 $p(\mathbf{x})$ は、 θ によってパラメータ化された条件付き分布 $p(\mathbf{x}|\mathbf{z};\theta)$ を用いて、

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

と書くことができる。VAE では、潜在変数の分布について、以下の 2 つの仮定を置く。

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

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

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。ベイズ統計では $p(\mathbf{z})$ は事前分布と呼ばれ、データ \mathbf{x} を観測した後の分布 $p(\mathbf{z}|\mathbf{x})$ は事後分布と呼ばれる。 $p(\mathbf{z}|\mathbf{x})$ を解析的に求めることができるケースは非常に限られており、これが解けない場合、近似分布 $q(\mathbf{z}|\mathbf{x})$ を導入して $p(\mathbf{z}|\mathbf{x})$ を近似することがベイズ統計ではよく行われる。VAE でも、 $p(\mathbf{z}|\mathbf{x})$ を別のパラメータ ϕ を用いて $q(\mathbf{z}|\mathbf{x};\phi)$ によって近似する。 $p(\mathbf{z}|\mathbf{x})$ はガウス分布を仮定しているため、 $q(\mathbf{z}|\mathbf{x};\phi)$ もガウス分布を仮定する。VAE では、この近似分布 $q(\mathbf{z}|\mathbf{x};\phi)$ もニューラルネットワークを用いて表現する。

VAE の目的はデータの分布 $p(\mathbf{x})$ を推定することであるため、目的関数は式 (2.1) の尤度を最大化することである。しかし、式 (2.1) は \mathbf{z} の周辺化を含み、これを解析的に求めることは困難であるため、尤度そのものを計算することはできない。そこで、式 (2.1) に先ほど定義した $p(\mathbf{z}|\mathbf{x})$ の近似分布 $q(\mathbf{z}|\mathbf{x};\phi)$ を導入し、その対数をとることで、以下のような変分下限を導出する。

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

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

ここで、式 (2.4) でイエンセンの不等式 (Jensen's inequality) を用いている。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 $p(\mathbf{z})$, $q(\mathbf{z}|\mathbf{x};\phi)$ 共にガウス分布を仮定して

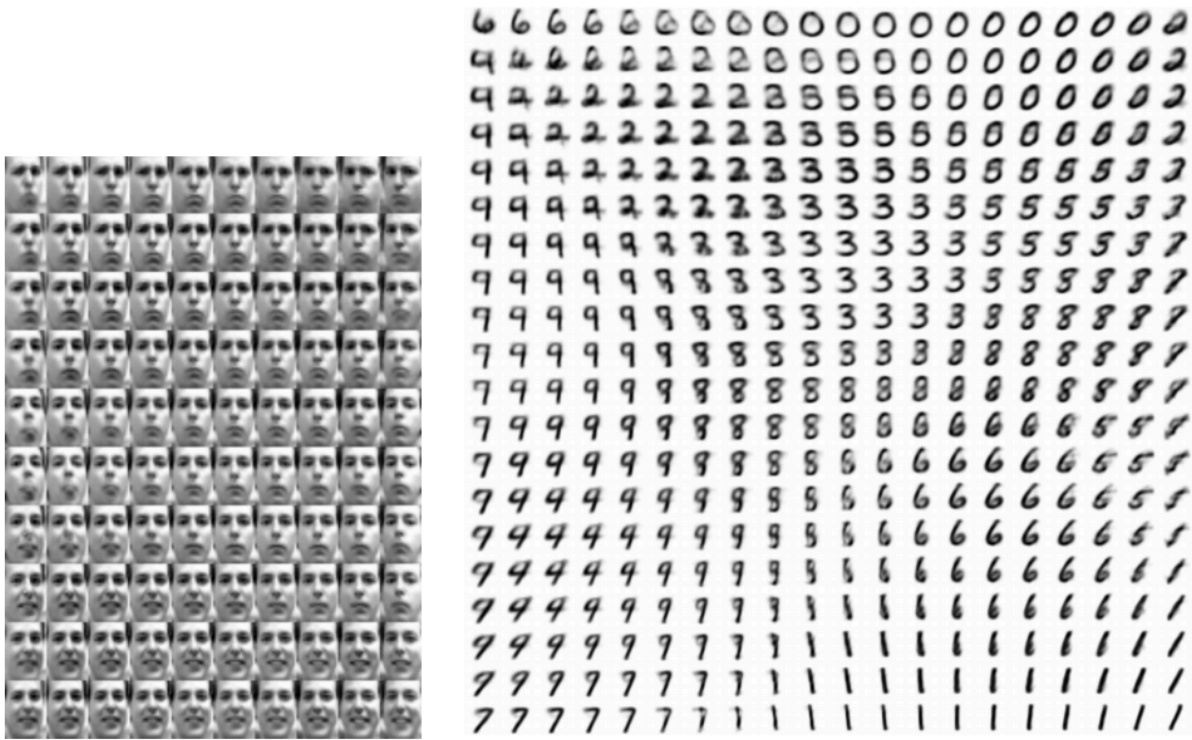


Fig. 2.2 VAE を用いて生成された画像の例 ([2] より引用)

いるため、解析的に計算することができる．一方、式 (2.5) の第 1 項は、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 $\frac{1}{L} \sum_l \log p(x|z)$ でモンテカルロ近似する．通常、 $p(x|z)$ はベルヌーイ分布や正規分布とし、その尤度を計算する．VAE ではこの変分下限を目的関数として最大化するように、誤差逆伝播法でニューラルネットワークのパラメータ θ と ϕ の最適化を行う．しかし、ここで $q(z|x; \phi)$ からサンプリングを行う部分で勾配の計算ができず、計算グラフが途切れてしまって、誤差逆伝播法を用いた最適化ができなくなってしまうという問題が生じる．これを解決するために、VAE では再パラメータ化トリック (reparameterization trick) を使用する．通常のサンプリングを行う場合、正規分布 $q(z|x; \phi)$ の母数である平均 $\mu(x)$ と標準偏差 $\sigma(x)$ がニューラルネットワークによって出力された後、 $N(\mu(x), \sigma(x))$ から z をサンプリングするが、これでは計算グラフが途切れてしまう．そこで、再パラメータ化トリックでは、 $N(\mu(x), \sigma(x))$ から直接サンプリングを行う代わりに、標準正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ からサンプルされる変数 ϵ を用いて、 $z = \mu(x) + \epsilon \cdot \sigma(x)$ と計算することによって、計算グラフを途切らせることなく、確率的なサンプリングを可能にする．

2.2 深層状態空間モデル (DSSM)

第 3 章

深層状態空間モデルの限界

3.1 学習が失敗した例

3.2 深層状態空間モデルの問題点

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 実験概要

adam

5.2 データセット

5.3 実験条件

5.4 評価指標

評価指標には，予測誤差 (負の対数尤度)，hogehoge スコアを用いる

5.4.1 hogehoge スコア

オリジナル/ノンオリジナル

5.5 学習時のテクニック

5.6 実験結果

5.6.1 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ふ
尤度が下がった

5.6.2 定量評価 (hoge hoge スコア)

5.6.3 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 ロボットマニピュレーション

6.3.2 物理シミュレーションの近似

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.