内的不確実性に基づく先読みが実現する適応的な動作学習

○昼間 彪吾（早稲田大学）　伊藤 洋（早稲田大学）　尾形 哲也（早稲田大学）

不確実な環境でロボットが状況を正しく認識し, 行動するには, 探索的な試行錯誤が必要である. しかし, 教示データをトレースする模倣学習ではその探索的な性質の獲得が難しい. 本提案では, 学習モデルが予測する内的不確実性に応じたランダムノイズをモデルの内部状態に加え, 先読みによる探索的な動作学習を実現する. 実験では提案手法が従来手法と比べて不確実性に対する適応性が高く, 効果的な動作学習が可能であることが示された。

1. 序論

深層学習モデルを用いたロボット制御技術の登場により, 近年ロボット適用が可能なタスク種類が増加している. 従来の制御手法では, ロボットが扱う事象を 1 つ 1 つルールとして書き出す必要があったため, 厳しく制約がかかったタスク空間にのみ適用可能であった. 一方深層学習を用いた学習ベースの手法では, 教示データを学習する過程で多様かつ複雑なルールを自動的に獲得できるため, より複雑で柔軟な制御が実現可能である.

深層学習を用いた代表的なロボット制御手法として模倣学習が挙げられる [1]. 模倣学習は, マニュアル操作による動作例を教示データとして模倣することで, 目的のタスク動作を学習させる手法である. 学習モデルは模倣を通じて, タスク実行時の感覚情報（例. カメラ画像やロボットアームの関節角度）の潜在的なダイナミクスが獲得する. そのため, 未知の状況下においても柔軟に目的のタスクを実行することが可能である. 故に, 人間の生活空間のように制約がかかっておらず, 自由度の高い環境への応用が期待されている.

ロボットを自由度の高い環境下で制御するにあたって, 環境の不確実性を考慮した制御技術が重要になってくる. なぜなら, ロボットが作用する対象物は常に状態が一定であるとは限らず, また客観的な観測からはその状態が推定できないことが多いからである. 例えば, 何気ないドア開け動作を取っても, それが押し戸か引き戸か見た目では判別できないといった状況が発生する. 故に, 環境に対して不確実性を仮定し, それに合わせて正しい動作を生成する適応的な制御技術が求められる.

しかし, 従来の模倣学習手法は不確実性が絡むタスクの学習を苦手とする場合が多い. これは, 同手法が事前に集めた教示データのみから学習しており, 予測モデルが直接環境と相互作用する経験を得られないからである. 模倣学習では主にタスク実行に成功した動作例のみから動作を学習するため, ドア開けタスクのように失敗したとき（例. 押戸を引いてしまったとき）に初めて判明する不確実性を正確に認識することが難しい. また, 動作中に不確実性の高い状況に遭遇した際, それを解決するためには探索的な行動を生成できる必要がある. しかし, 同様に模倣学習では実際の環境で探索的に行動した経験を与えることができないため, そのような動作も生成されにくい.

以上より, 不確実性の高いタスクを学習するためには, 正確に環境の不確実性を理解することと, 探索的に行動することで不確実性を解消しつつタスク動作を実現することが求められる. 本稿は, 従来の模倣学習モデルを拡張し, 不確実性の高いタスク環境下で適応的な動作が実現可能な学習モデルを提案する. 提案モデルでは, ロボット動作予測モデルに先読み機能を追加し, タスク実行時の不確実性の長期的な影響を考慮した動作学習を実現する. この先読み機能では, モデルの内部状態にランダムノイズを加えた上で未来の状態を脳内探索により複数パターンシミュレートする. その探索結果から将来の最も不確実性が低くなるパターンを推定し, 動作生成に反映することで適応的な行動を生成する.

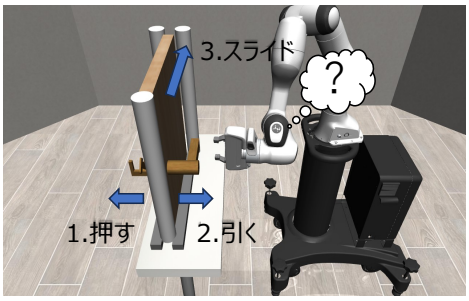


図 1　不確実性を含むロボットタスクの例（ドア開け）

2. 関連研究

2.1 不確実性の予測

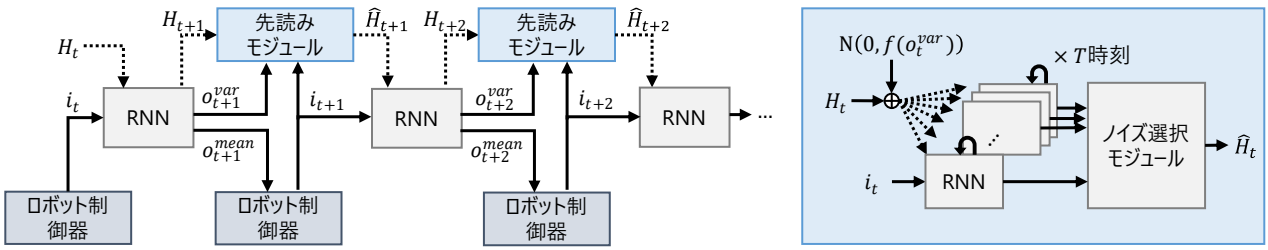
模倣学習では, 感覚情報の時系列データを学習するために Recurrent Neural Network （RNN）がしばしば用いられる [1]. RNN は単位時刻毎に予測を行い, 各予測の間で内部状態を引き継ぐことで時間的なダイナミクスの学習を実現する. しかし, 同モデルは決定論的な構造を持っているため, 確率的な挙動を必要とする不確実なタスクの学習は苦手である.

RNN に確率的な性質を付与するため, 村田らは RNN の予測に分布を仮定した Stochastic RNN（SRNN）を提案した [2]. SRNN では, 感覚情報の平均と分散を予測し, それぞれ真値との予測誤差の計算と, 学習への影響度を決定する重みとして学習に使用する. 一般的に不確実性の高いデータほど学習が難しく, 全体の予測誤差が下がりにくい原因となるため, SRNN ではそれらの影響度が下がるようにモデルが最適化される. そのため, 影響度を決定する分散値には, 対象の感覚情報に想定する不確実性が反映された値とであると解釈できる.

上記構造により SRNN は環境の不確実性を捉えることが可能だが, 動作予測時は依然 RNN の決定論的な性質が残る. 故に探索的な動作は発生しにくく, 不確実性を持つタスクの学習が難しい.

2.2 探索的な動作予測

適応的な動作予測をする方法として, フリストンら [3] が提案した能動的推論というフレームワークがある.



(A) 提案モデルのネットワーク構造　　(B) 先読みモジュールの構造

図 2　先読みモジュールを導入したロボット動作学習モデルの構造

能動的推論では, 外界に関する観測情報と, エージェント内の外界に関する生成モデルの予測を比較し, 予測誤差を最小化するように推論を行う. 予測誤差を最小化する方法は大きく分けて 2 つあり, 観測に合わせてエージェントの信念を更新する方法と, エージェントの信念に沿う観測が得られるように行動を調整する方法である. 同フレームワークをロボットの動作予測に用いた場合, 前者は観測に合わせて予測モデルの内部状態を更新すること, 後者は対象物の不確実性に基づいて探索的な行動を生成することに該当する.

先行研究 [4, 5] では, 能動的推論を簡易的なシミュレータ実験に適用し, 不確実なタスクに対して探索的な行動生成を実現した. しかし, 同研究では不確実性のモデルが既知であることや, 探索する行動空間が小さいなどの制約を前提としていた. ロボットタスクは両方の制約を満たす場合は少ないため, 現実的な手法とは言えない. また, 谷ら [6] は RNN ベースの予測モデルに能動的推論の機能を組み込むことで, ロボットのような複雑なタスク空間への応用を実現した. しかし, 同モデルは誤差逆伝搬法による内部状態の更新を毎タイムステップ実施するため動作生成速度が遅く, 現実的なロボット応用が難しい. したがって, 探索的な動作を生成可能かつ計算コストのオーバーヘッドが少ない動作生成手法が求められる.

3.　手法

本研究の提案手法は, 深層予測学習のフレームワークをベースとしたロボット動作予測モデルである. 図 2(A) が示すように, 入力は時刻 t の感覚情報 i_t であり, 出力は次の時刻 $t+1$ で期待される感覚情報 o_{t+1} である. 特に入出力の連続する感覚情報の関係性の学習には SRNN[2] を用いたため, 各感覚情報に関して平均値 o_{t+1}^{mean} と分散値 o_{t+1}^{var} が予測される. 動作生成時は予測された平均値をロボット制御器に入力し, その後取得したデータをもとに再度予測を行う. この処理を繰り返し実行することで, 特定のタスク動作が生成される. 学習時は [2] 同様, 予測した感覚情報の尤度を最大化するようにモデルを最適化した.

感覚情報は, 環境を撮影したカメラ画像と, ロボットアームの関節角度情報で構成されている. 前者は Convolutional Auto-Encoder (CAE) により次元圧縮を行い, RGB 画像からベクトル表現に変換されたデータを RNN の入出力に用いる.

本提案モデルが導入した先読みモジュールの構造を図 2(B) に示す. 先読みモジュールでは, RNN の内部状態 H_t を脳内シミュレーションを通じて将来の予測分散が下がるような内部状態 $\hat{H}_t$ に更新する. ここで, 脳内シミュレーションは, Closed-loop 予測という手法により実現される. Closed-loop 予測とは, RNN において自身の出力値を次の時刻の疑似入力値として用いることで, 外界の観測データなしに未来の状態を予測する手法である. これにより, RNN が学習の過程で獲得した感覚情報のダイナミクスと, 現時刻の RNN の内部状態 (=信念) を基にした先読みが可能になる. この先読みは学習時と動作生成時の両方で行われる.

先読みの具体的な処理としては, まず H_t をベースとした n パターンの内部状態を基に, それぞれ T ステップ先まで先読みを行う. 異なるパターンの内部状態は, H_t に対して $\mathcal{N}(0, f(o_t^{var}))$ からサンプルしたランダムノイズを付与することで生成される. $f(\cdot)$ は入力値を $[0.05, 0.15]$ に正規化する関数である. これにより異なる内部状態を出発点とした先読みが実施され, T ステップ先の予測平均 $o_{t+T}^{n,mean}$ と予測分散 $o_{t+T}^{n,var}$ に差異が生じる. 先読み結果はノイズ選択モジュールに入力され, 最も小さい $o_{t+T}^{n,var}$ を出力したノイズが選択される. そのノイズは H_t に付与され, 本番予測時に用いられる.

また, 各時刻で付与されるランダムノイズの強度は前回の予測分散値 o_t^{var} に比例するように設計されている. そのため, 前ステップで予測分散が高く, 不確実性が高いと認識された場合はより強いノイズを付与することで探索的な先読みが行われる. 逆に不確実性が低い場合はノイズ強度が減り, 探索的な挙動が抑制される. 以上により, 提案モデルは目的のタスク環境に対して不確実性を予測しつつ, その度合いに応じて行動の探索性を決定する, 適応的な動作が実現可能となる.

4.　実験

実験では, 不確実な環境下でのモデルの適応性を検証するため, 図 1 に示すドア開けタスクの学習を行った. タスクで使用したドアは, ドアノブを捻ったあと, 押す, 引く, スライドのいずれかの動作で開く. タスク実行時は, これら 3 種類のドアからランダムに 1 つ提示される. 各ドア種は見た目では判別できず, 実際にドアを動かしたときの挙動をもとに種類を推定する必要がある.

学習データは Robosuite[7] というシミュレータ環境で収集した. ロボットはグリッパーを含む 7 自由度のアームロボット使用しており, 手先制御で動作教示を行った. 学習データは $128 \times 128 \times 3$ の RGB カメラ画像とアームの関節角度情報で構成される時系列デー

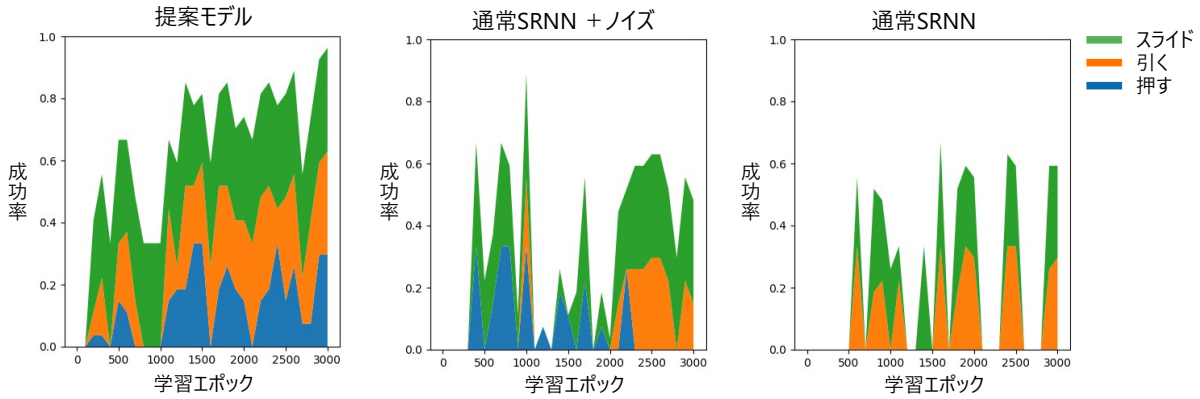


図 3　ドア開け動作の成功率

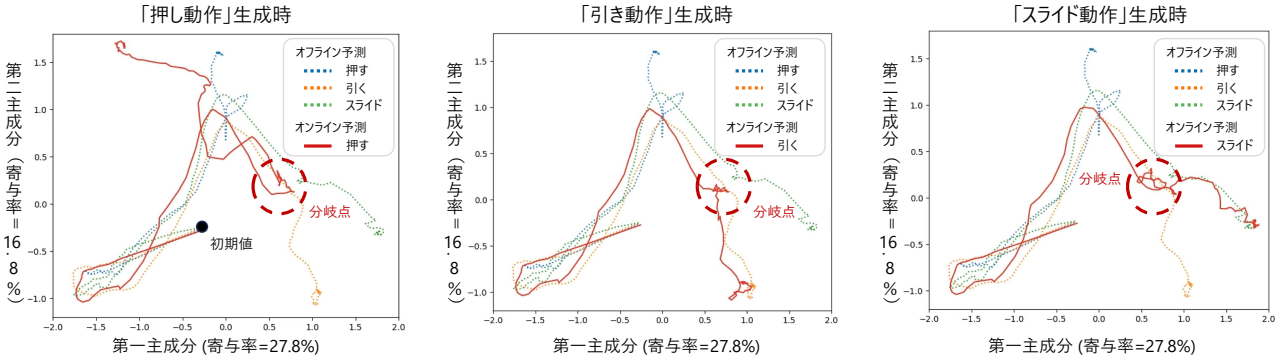


図 4　提案モデルで動作生成した際の内部状態の遷移

タであり，ドア種毎に 5 個，計 15 個のデータを収集した．各動作例は 100 時刻分の長さを持ち，動作教示時は 10Hz でデータを記録した．

モデルは 3000 エポック学習しており，最適化手法には Adam を使用した．また，先読みモジュールでは $n = 5$ パターンのノイズに対して，それぞれ $T = 10$ 時刻分の先読みを予測し，内部状態の更新を行った．

また，比較実験では従来手法として，先読みモジュールを持たない通常の SRNN モデル (=通常 SRNN モデル) と，同モデルの内部状態に毎時刻ランダムノイズを加えた派生モデル (=通常 SRNN+ノイズモデル) を使用した．特に後者は，ランダムノイズの付与が探索的な動作を実現する上で重要である一方で，いかに先読み機能が重要な役割を持つかを検証するために用いた．

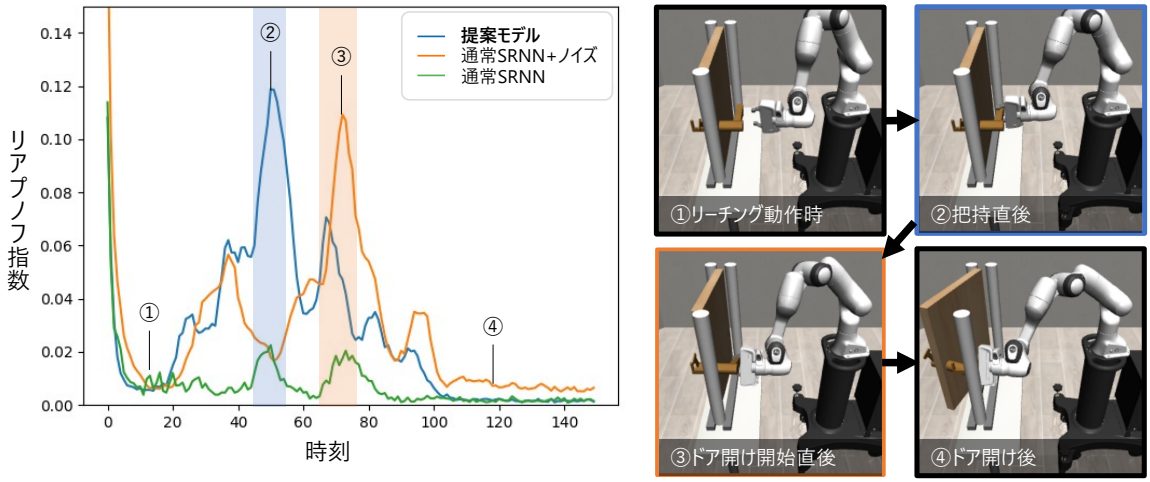
5.　結果

まず，各予測モデルが不確実性を含むドア開けタスクに対して正しく学習できたかを検証するため，ドア開け動作の成功率を評価した．評価時はドアやロボットの初期位置をランダムに動かしつつ，ドア種毎に 10 回ずつ動作生成を行った．図 3 に各モデルの成功率を，ドアの種類ごとに色分けして示す．図が示すように，提案モデルは学習初期から 3 種類の動作それぞれの動作生成に成功しており，1300 エポック以降は平均して 80%の確率で各種ドアを開けることに成功した．

図 4 では，提案モデルで各ドア種を開ける動作を生成した際の RNN の内部状態の時間遷移をプロットしている．点線は各ドア種の学習データでオフライン予測した場合の遷移経路（アトラクター）であり，実線はオンラインで動作生成した際の遷移経路である．図が示すように，提案モデルの内部状態はオンライン予測中に各ドア種のアトラクター間を行き来するような挙動を示した．その結果，自己組織化された分岐点で探索的な行動が発生し，最終的に正しい動作のアトラクターに引き込まれる結果となった．またこの分岐点では，他の点と比べてノイズを加えた際に様々に行動が変化する傾向が高いため，カオス的な性質を持つアトラクターであるといえる．以上より，提案モデルは学習を通して正しく環境の特性を理解し，適応的な動作を実現する RNN の内部状態空間を構築できたといえる．

一方，従来手法は学習した 3000 エポックの間，基本的に 3 種類あるドアのうち，同時に 2 種類までしか正しい動作を生成できず，結果低い成功率を示した．RNN の内部状態の遷移を解析したところ，通常 SRNN モデルは明確な分岐点の獲得に失敗しており，通常 SRNN+ノイズモデルは分岐点は発生しているものの，正しいアトラクターへの遷移に失敗していることが分かった．

学習モデルの成功率の差は，各モデルが不確実性 RNN の内部状態空間にどのように埋め込んだかによって発生したと考えられる．図 5 は，押戸の学習データを用いてオフライン予測した場合の，各予測モデルの RNN 内部状態のリアプノフ指数を比較している．リアプノフ指数とは状態のカオス性を示す指標であり，この場合は各時刻の RNN の内部状態がどの程度発散するように変化し得るかを定量的に示す値である．カオス性が高いほど様々な状態へ遷移する可能性を同時に持つこと



(A) 各学習モデルにおける不確実性の埋め込み　　(B) 各時点でのロボットの状態

図 5　押戸のドア開け動作において各時点で想定された不確実性の比較

を示しており, 各時点で想定している内的な不確実性を表す値と解釈できる.

通常 SRNN モデルが一貫して低い値を示していたのに対して, 提案モデルと通常 SRNN+ノイズモデルは, タスク動作の特定のタイミングで強いピーク示すように学習されていた. 前者はドアノブを把持したタイミング, 後者はドア開け動作が開始した直後であった. それぞれ, 不確実性のあるドアの開き方が決定する事象に対して, 原因となる行動方策と, その結果となる観測に不確実性が埋め込まれたと考えられる. SRNN モデルでは, 予測しにくい感覚情報に対して予測分散が高くなるように学習するため, 複数パターン存在するドア開け直後の観測に高い不確実性が埋め込まれるのは自然な結果である. しかし, 不確実性の理解を観測結果に依存させた場合, 不確実性を埋め込んだ特定の状態が得られない限り, 推論時に不確実性が正しく表現されない. そしてその不確実性が予測動作の探索性を決定するため, 目的の動作を達成する適応的な動作が生成されにくいという結果に繋がる.

このような受動的な表現に対して, 提案モデルは能動的な表現を獲得した. 提案モデルは, 不確実性を行動方策に紐づけて学習したため, 観測結果に関わらず高い不確実性を想定した動作生成が行われる. これは, 複数の先読み結果の予測分散を考慮した行動選択を行ったことで, より長期的な影響を考慮した不確実性が構造化されたからだと考えられる. そのため, 誤った動作のアトラクターに引き込まれたとしても, 想定した感覚情報が得られない場合は不確実性の高いカオスアトラクターに戻り, 再度探索的に正しいアトラクターへの遷移を模索することが可能となる. これにより状況に応じた適応的な動作が生成可能になり, 高い成功率に繋がったのだと考えられる.

6.　まとめと今後の展望

本研究は不確実なロボットタスクの学習を実現するため, 先読みモジュールを付与した深層予測学習モデルを提案した. 提案モデルは, 毎ステップの予測にて脳内シミュレーションを行い, 複数サンプリングした未来のうち予測分散が最も下がる内部状態を選択することで, 適応的な動作の学習を実現した. 同手法は軽量且つ高速であるため, より高頻度で試行錯誤を必要とする動的な環境への適用が期待される. 今後は, より動的で複雑性の高い実機環境への適用を目指す.

参　考　文　献

[1] Hiroshi Ito, Kenjiro Yamamoto, Hiroki Mori, and Tetsuya Ogata. Efficient multitask learning with an embodied predictive model for door opening and entry with whole-body control. *Science Robotics*, Vol. 7, No. 65, p. eaax8177, April 2022.

[2] Shingo Murata, Jun Namikawa, Hiroaki Arie, Shigeki Sugano, and Jun Tani. Learning to reproduce fluctuating time series by inferring their time-dependent stochastic properties: Application in robot learning via tutoring. *IEEE Transactions on Autonomous Mental Development*, Vol. 5, No. 4, pp. 298–310, 2013.

[3] Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *Journal of Physiology-Paris*, Vol. 100, No. 1, pp. 70–87, 2006. Theoretical and Computational Neuroscience: Understanding Brain Functions.

[4] Kai Ueltzhöffer. Deep active inference. *Biological Cybernetics*, Vol. 112, pp. 547–573, 2018.

[5] Philipp Schwartenbeck, Johannes Passecker, Tobias U Hauser, Thomas HB FitzGerald, Martin Kronbichler, and Karl J Friston. Computational mechanisms of curiosity and goal-directed exploration. *eLife*, Vol. 8, p. e41703, may 2019.

[6] Ahmadreza Ahmadi and Jun Tani. A Novel Predictive-Coding-Inspired Variational RNN Model for Online Prediction and Recognition. *Neural Computation*, Vol. 31, No. 11, pp. 2025–2074, 11 2019.

[7] Yuke Zhu, Josiah Wong, Ajay Mandlekar, Roberto Martín-Martín, Abhishek Joshi, Soroush Nasiriany, and Yifeng Zhu. robosuite: A modular simulation framework and benchmark for robot learning. In *arXiv preprint arXiv:2009.12293*, 2020.