

信頼される AI の基盤技術  
2020 年度採択研究代表者

2022 年度  
年次報告書

小林 泰介

情報・システム研究機構 国立情報学研究所  
助教

頑健性と安全性の性能限界を明らかにする深層強化学習

## 研究成果の概要

本研究では深層強化学習が獲得する制御器を信頼できるものとするべく、その安全性および頑健性の評価・最大化を目指している。2022年度では、i) スパースな低次元空間上でのモデル学習、ii) 実時間での安全なプランニングを達成するためのモデル予測制御、の2点について研究開発に取り組んだ。

i)として、開発する制御器の計算コストは観測次元の大きさに強く影響されるため、制御に必要な最小限の次元を抽出し、そこでのモデルを学習することで制御のリアルタイム性向上および数値最適化の収束性向上を図った。具体的には、低次元空間の抽出技術として知られる変分オートエンコーダをツァリス統計に基づき再定式化し、その挙動を解析することで不要な次元が常に0を取るスパース化の条件を明らかにした。スパース化された次元をマスク処理で除外した空間上でのモデルを学習したところ、モデルの予測精度を落とすことなく予測にかかる計算コストの大幅な削減に成功し、実ロボットのリーチングタスクでも動作の高速化を達成した。

ii)として、2021年度に導出した、逆方向のKullback-Leiblerダイバージェンスの最小化問題に基づくモデル予測制御に対して、さらなる改良を施した。具体的には、従来はサンプルされた行動系列候補に含まれる正例を優先する更新と負例を除外する更新が干渉し合う問題が残されていたため、棄却サンプリングを参考にして、正例・負例を生成する提案分布を別々に更新した後に、それらを適切に合成した分布から行動系列候補をサンプリングするアルゴリズムを開発した。また、ここでの更新が鏡像降下法で為される点に注目し、Nesterovの加速法を更新の不安定さを加味しながら統合した、新しい更新アルゴリズムを開発した。これらを組み合わせることで、計算資源がCPUのみに限定されるような実ロボットに対しても、リアルタイムでの動作最適化に成功した。

### 【代表的な原著論文情報】

- 1) “Design of Restricted Normalizing Flow towards Arbitrary Stochastic Policy with Computational Efficiency”, *Advanced Robotics*, 2023 (in press)
- 2) “Sparse Representation Learning with Modified q-VAE towards Minimal Realization of World Model”, *Advanced Robotics*, 2023 (accepted)
- 3) “サンプリングベースモデル予測制御における棄却サンプリングの検証”, 第40回日本ロボット学会学術講演会, 1F1-06, 2022