

信頼される AI の基盤技術  
2020 年度採択研究者

2021 年度 年次報告書
------------------

小林 泰介

奈良先端科学技術大学院大学先端科学技術研究科  
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

## § 1. 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。2021年度では、基盤となるモデルベース強化学習において必要とされる、i) 確率的モデル学習における学習規範のメタ最適化、ii) 学習したモデルを用いた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、iii) より表現力の高い確率的モデルの設計手法、3点について研究開発に取り組んだ。

i)として、2020年度の成果である、バイアス・バリエーションのトレードオフを陽に考慮可能な学習則に対して、そのトレードオフのバランスをどのように取るのかという学習規範を単一のハイパーパラメータで表現される事実に注目した。このハイパーパラメータを方策勾配法を参考に確率的にメタ最適化する手法を確立することで、モデルの学習プロトコルと同時にメタ最適化を可能とした。

ii)として、最も汎用性が高いモデル予測制御手法であるクロスエントロピー法が順方向の Kullback-Leibler ダイバージェンスの最小化問題を解いている点に注目した。準最適解でも良いので早期に安全な解を見出させるよう、逆方向の Kullback-Leibler ダイバージェンスに関する最小化問題へと変換し、それを解くアルゴリズムを導出した。

iii)として、多様な確率分布形状を表現可能な Normalizing Flow モデルに注目した。このモデルの課題であった、その統計量を閉形式で効率良く算出できない点を改善してモデルの平均的な挙動を推測できるようにすべく、設計上の制約条件を導出した。また、その制約によって損なわれる表現力を再度得るための設計指針も明らかにした。

### 【代表的な原著論文情報】

- 1) “Meta-Optimization of Bias-Variance Trade-off in Stochastic Model Learning”, IEEE Access, Vol. 9, pp.148783-148799, 2021
- 2) “カルバック・ライブラ情報量の非対称性に着目したサンプリングベースモデル予測制御”, 日本ロボット学会誌, Vol. 40, No. 2, pp.174-177, 2022
- 3) “制約付き Normalizing Flow と強化学習への応用”, 第 27 回ロボティクスシンポジウム, pp.129-132, 2022