

2023 年度年次報告書
信頼される AI の基盤技術
2022 年度採択研究代表者

和賀 正樹

京都大学 大学院情報学研究科
助教

品質保証と説明の両立による信頼できる AI の構築技術

研究成果の概要

まず本年度は、AI システムにおいて重要となる様々な性質に関する形式検証の適用可能性を向上するため、パラメタ付き時間オートマトンに対するハイパープロパティの形式検証手法を構築した。形式検証の分野では従来単一の実行列に対する性質の検証が広く行われてきたが、より一般の性質を扱うため、複数の実行列間の関係に対する性質を記述できるハイパープロパティに対する検証も近年注目されている。AI システムにおいても、例えば頑健性や公平性など複数の実行列間の関係に関する性質は重要とされている。本研究では具体的にはパラメタ付き時間オートマトンという時間制約に関するパラメタを含むことのできる時間オートマトンによって記述されたシステムと、本研究において提案した HyperPTCTL という論理によって記述した仕様について、システムが仕様を充たすようなパラメタの値の範囲を導出する手法を提案した。本研究の成果を用いることで、例えばある程度ブラックボックスな AI コンポーネントを含む制御システムが、ある種の公平性を充たすために要求される内部の時間制約を導出することができる。本研究の成果を論文としてまとめ、現在査読中である。

さらに昨年度から取り組んでいたブラックボックスかつ確率的な挙動をする AI システムのための手法について、今年度は収束性の保証など主に理論的な結果を挙げた。昨年度の成果と含め、これらの結果をまとめた論文は組み込みシステムにおけるトップ国際会議 EMSOFT 2023 に採択され、ACM Transactions on Embedded Computing Systems に採録された。

【代表的な原著論文情報】

- 1) Shijubo, Junya, Masaki Waga, and Kohei Suenaga. "Probabilistic Black-Box Checking via Active MDP Learning." ACM Transactions on Embedded Computing Systems 22.5s (2023): 1-26.