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

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

本年度はまず、品質保証と説明の両立による信頼できる AI の構築技術のための基盤技術として、オートマトン学習アルゴリズムに取り組んだ。具体的には、決定的オートマトンに対する能動的学習アルゴリズムである L*アルゴリズムを、連続的時間概念を持つような決定的時間オートマトンに対して拡張した。本研究では、正規言語の Myhill-Nerode の定理による特徴付けと類似した特徴付けを決定的時間オートマトンの認識する言語クラスに対して与え、本特徴付けに基いて学習アルゴリズムを構築した。本手法を用いることで、AI システムの中でも特に実時間システムや物理情報システムのような時間に対する制約が重要となるシステムの説明性を向上させられることが期待される。本研究の成果をまとめた論文を執筆し、システム検証のトップ国際会議である CAV 2023 に採択された。

また、確率的な挙動をする AI システムのための信頼性向上手法として、マルコフ決定過程 (MDP) の学習と確率的モデル検査を用いたテスト手法の構築も行った。具体的には、AI システムの挙動を近似する MDP を能動的 MDP 学習アルゴリズムを用いて自動で学習し、その MDP に対して確率的モデル検査による形式検証を行うことで、品質の保証と挙動の説明を両立できるテスト手法を構築した。本研究の成果を論文としてまとめ、現在査読中である。

【代表的な原著論文情報】

1) Waga, M. "Active Learning of Deterministic Timed Automata with Myhill-Nerode Style Characterization." To appear in Proc. CAV 2023.