

2024 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

本年度はまず、品質保証と説明性の両立によって信頼性の高い AI システムを構築する基盤技術として、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、記号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 Λ^* アルゴリズムを、連続的時間概念を持つような記号的ミリーオートマトンに対して拡張した。本研究では、本質的に重要な入力集合を推定しながら学習を進めることで、ミリーオートマトンに必要な出力関数の学習や計算量の解析などを行った。本手法を用いることで、数値など無限通りの入力を受け取る AI システムの説明性を向上させることが期待される。本研究の成果を論文としてまとめ、現在査読中である。

また、AI システムにおいて重要な頑健性や公平性などのハイパープロパティを対象とする実行時モニタリング手法を構築した。本研究では従来の正規言語を拡張した、非同期正規ハイパー言語という言語クラスを提案し、その言語の上でのパターンマッチングアルゴリズムを提案した。本研究の成果を用いることで、例えばログ中の複数箇所を比較しながら、公平性の違反の証拠となる部分の集合を抽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。

昨年度から取り組んでいたパラメタ付き時間オートマトン向けハイパープロパティ形式検証手法に対し、追加の実験評価を行った。これらの結果をまとめた論文は組み込みシステム分野の国際会議 EMSOFT 2024 に採択され、さらに IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems に採録された。

【代表的な原著論文情報】

- 1) Masaki Waga and Étienne André. "Hyper parametric timed CTL." IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 43.11 (2024): 4286-4297.