



HOKKAIDO UNIVERSITY

Title	Rule-Based Machine Learning and Abductive Reasoning for Explainable AI
Author(s)	佐々木, 耀一
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16671号
Issue Date	2025-09-25
DOI	https://doi.org/10.14943/doctoral.k16671
Doc URL	https://hdl.handle.net/2115/98281
Type	doctoral thesis
File Information	Yoichi_Sasaki_review.pdf, 審査の要旨 https://creativecommons.org/licenses/by/4.0/



学位論文審査の要旨

博士の専攻分野の名称 博士 (情報科学) 氏名 佐々木 耀一

審査担当者 主 査 教 授 有村 博紀

 副 査 教 授 吉岡 真治

 副 査 教 授 堀山 貴史

学位論文題名

Rule-Based Machine Learning and Abductive Reasoning for Explainable AI

(説明可能 AI のためのルールベース機械学習とアブダクティブ推論)

本論文は、ルールベースの機械学習システム運用における高精度な予測と、予測結果の解釈性を両立させるための説明可能機械学習の諸問題について考察し、ブラックボックスモデルとルールベースを融合した新しいルールベースの機械学習の枠組みを提案したものであり、離散最適化ソルバーを用いた効率良い実現方法を開発し、その性能と有用性を検証したものである。

深層学習器やアンサンブル学習器に代表される高精度なブラックボックスモデルの利用増加と、その産業や社会における受容の急速な拡大により、機械学習システムの利用者や運用者に対する予測結果の解釈や説明を提供するための「説明可能 AI」(explainable artificial intelligence, XAI) の研究への関心が高まっている。例えば、医療診断や与信審査では、予測結果の根拠やモデルの修正方法を提供することで、システムが提供するサービスの品質や、システム運用の安定性、顧客の満足度向上を期待できる。

説明可能機械学習の文脈では、記号的 AI とデータ駆動型 AI の融合も重要である。本研究では、ルールを共通の表現とすることで両者を接続し、ユーザーが説明結果に納得できる説明可能な AI を構築することを目的とする。これにより、データからのルールベースの説明が可能になり、さらに、ルールとユーザの背景知識の整合性検証が可能になる。そこで、この枠組みに必要なルールベース XAI の問題について、データ駆動的 AI (第 4 章, 第 5 章) と記号的 AI (第 6 章) の両方の観点から検討する。

第 4 章では、代理ルールモデル (alternative rule model, ARM) と呼ばれる新しい機械学習モデルを提案する。本章では、与えられたブラックボックスモデルの予測結果との誤差と、与えられたルール集合から説明として選択するルール数を最小化する組合せ最適化問題を考え、これをホーン節のみで構成される重み付き MaxSAT 問題へ帰着することで、近年の高速な SAT ソルバーを用いて実用的な速度で解く方法を提案する。

第 5 章では、新しいルールベースモデルとして、ルールリストを一般化し、入力に対して条件を満たす最初の 1 つのルールではなく、 k 個のルールの予測を統合することで予測を行う Top- k ルールリストを提案する。モデルの学習問題を整数線形計画 (ILP) 問題として定式化し、効率的に学習する新しい手法を提案する。

第 6 章では、アブダクションの一種である重み付きアブダクションの効率良い実現方法を考察し、する。アブダクションとは、背景知識を仮定し、与えられた観測を説明する仮説を求める XAI 技術の一つであるが、その効率良い実現方法は課題であった。そこで、これを制限された論理式である

ホーン節を用いた重み付き名大論理式充足問題 (重み付き MaxSAT) 問題として定式化し, 重み付き MaxSAT ソルバーを用いた新しい計算手法を提案し, 計算効率向上を実証的に評価した.

これを要するに, 本論文は, 人工知能における機械学習システムの運用問題に関して, 高い予測精度と良好な説明可能性を両立させる新しいルールベース機械学習手法の設計と実装が可能であるという新知見を得たものであり, 情報科学におけるデータ指向の人工知能技術分野において貢献するところ大なるものがある. よって著者は, 北海道大学 博士 (情報科学) の学位を授与される資格あるものと認める.