



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_abstract.pdf, 論文内容の要旨 https://creativecommons.org/licenses/by/4.0/



学 位 論 文 内 容 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

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）

ディープニューラルネットワークや勾配ブースティングなどを含むブラックボックスモデルの発展に伴い、様々なタスクにおいて AI による高精度な予測が可能になってきている。一方で、医療診断や与信審査などを含む実世界の様々な問題に AI を適用する際には、AI の予測を人間が検証し判断する必要があることから、説明可能 AI (explainable AI, XAI) に関する研究は、近年大きな注目を浴びている。

人工知能分野での研究の初期は、論理に基づく記号的 AI と呼ばれるアプローチが主流であった。なかでもアブダクションは、観測結果に対して最も適切な説明や仮説を導くための推論手法として、初期から注目されてきた。論理式の集合で表現される背景知識と観測が与えられたとき、アブダクションは、観測を説明するための論理式を仮説として導出する方法である。この仮説を背景知識に追加したとき、観測が論理的に導出され、かつ矛盾が生じない必要がある。近年では、XAI の文脈において再び注目を集めており、その理由は、説明に用いる知識が論理式で表現されていることから、ルールとしてユーザーにとって理解可能であり、出力もそのルールを用いて厳密な論理に基づいて導出されることから説明に曖昧性が存在しない点にある。一方で、アブダクションには計算困難性、知識獲得のボトルネック、柔軟性の欠如といった課題も知られている。

記号的 AI と並行して、データ駆動型 AI も広く研究されてきた。こちらは、データから自動的に入力と出力の関係を学習できるという利点がある。XAI の文脈では、単純で解釈可能なモデルを使う手法と、ブラックボックスモデルに対する後付け説明を生成する手法に大別される。前者の手法では、線形モデル、決定木、決定リストなどの単純で解釈可能なモデルを用いて、予測精度を保ちつつ説明性を保持することが目指される。中でも、ルールベースモデルは、人間にとって直感的であり、高い解釈性を持つとされる。一方で、これらのモデルはブラックボックスモデルに比べて予測精度が劣る場合があるのが課題である。後者の手法では、任意のブラックボックスモデルの個々の予測に対して局所的な説明を生成する。しかし、入力ごとに説明を生成する必要があり、人間がすべての説明を検証するにはコストがかかる。

これまでデータ駆動型 AI と記号的 AI による説明生成の研究の多くは個別に行われてきたが、得られた説明に対してユーザーが納得するためには、データやブラックボックスモデルから得られる新しい知見と、ユーザーの既存知識に基づく論理的解釈を結びつける必要がある。

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。その結果、データから単にルールベースの説明を得るだけでなく、それらに論理的矛盾がないか、より良い説明がないかをユーザーの知識と組み合わせてフィードバックすることができ、一貫通貫に検証できる。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、データ駆動型 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。このモデルは、入力データに対し、ルール集合の中で条件を満たしつつ、内包するブラックボックスモデルが出す予測と最も近い予測を出す決定ルールを選択し、そのルールに基づいて予測を出力する。選択されるルールの予測値が十分ブラックボックスの予測値と近ければ、ARM 自体がブラックボックスとほぼ同様の予測精度を保持したまま常に根拠付きで説明できる検証可能なモデルだといえる。ARM の学習時には、ブラックボックスモデルの予測結果との誤差を最小にしつつ、与えられたルール集合の中から実際に説明として選択できるルールの候補数を同時に最小化することを考える。また、この組合せ最適化問題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 Top- k ルールリストを提案する。これは従来のルールリストを一般化したものであり、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 個のルールを使用し、それらの予測をアンサンブルすることで予測を行う。このモデルのパラメータ k を変更することにより、解釈性と精度のトレードオフが実現可能になる。このモデルの学習問題を、ルールの確信度スコアに基づいて事前に順序付けることで、整数線形計画 (ILP) 問題で定式化し、効率的に学習できる手法を提案する。

第6章では、アブダクションの代表的なモデルの一つである重み付きアブダクションを考える。一般にアブダクションは NP 困難であり、ILP や解集合プログラミングといった手法では、大規模な問題を現実的な時間で解くことが難しいという問題がある。そこで、深さ制限付き重み付きアブダクションを重み付き MaxSAT 問題として定式化する新しい手法を提案する。この定式化で得られる制約の多くがホーン節で構成されることを解析し、最新の SAT ソルバーの恩恵を受けやすいことを、計算機実験を通して確認する。

本研究を通して、人間の意思決定を支援することを目的としたルールベース XAI の複数の問題に取り組み、適切な離散最適化問題に定式化することで解決した。本研究の成果は、複数の観点から人間が検証可能な説明を与える枠組みであり、学術的にも産業的にも、人間が納得して意思決定を行う AI 社会の推進に貢献し得るものである。