

氏 名 竹村 彰浩

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 2365 号

学位授与の日付 2022 年 9 月 28 日

学位授与の要件 複合科学研究科 情報学専攻  
学位規則第6条第1項該当

学位論文題目 Knowledge Representation Methods for Bridging Machine  
Learning and Logical Reasoning

論文審査委員 主 査 井上 克巳  
情報学専攻 教授  
山田 誠二  
情報学専攻 教授  
杉山 磨人  
情報学専攻 准教授  
市瀬 龍太郎  
東京工業大学 工学院 教授  
佐藤 泰介  
国立情報学研究所 特任研究員/  
東京工業大学 名誉教授

## 博士論文の要旨

氏名 竹村 彰浩

論文題目 Knowledge Representation Methods for Bridging Machine Learning and Logical Reasoning

Recent advancements in machine learning (ML) and deep learning have led to a wide range of artificial intelligence (AI) applications. Meanwhile, knowledge representation and reasoning (KR) has continued to play a central role in the AI research by advancing our understanding of explicit representation of knowledge and inference. The integration of high-level symbolic reasoning with efficient low-level perception is one of the longstanding challenges in the field of AI. Integration of KR and ML is a promising avenue for research and may hold the key to solving some challenges in the respective subfields of artificial intelligence. A true integration of these two major fields of research is well beyond the scope of this thesis. Nonetheless, this thesis explores various approaches and techniques for the integration of machine learning and knowledge representation.

The first contribution of this thesis, presented in Chapter 3, involves the development and evaluation of a method to explain trained machine learning models. More specifically, a novel application of Answer Set Programming (ASP) is presented where the goal is to generate rule sets as explanations for trained tree-ensemble models. ASP is a declarative programming paradigm for solving difficult search problems. This problem is of practical relevance in business domains where understanding of the behavior of high-performing machine learning models and extraction of knowledge in human-readable form can aid users in the decision-making process. An advantage of using ASP is its expressiveness and extensibility, especially when representing various constraints. To our knowledge, ASP has never been used in the context of rule set generation from tree-ensembles.

To demonstrate the adaptability of our approach, we present implementations for both global and local explanations of learned tree-ensemble models using our method. In general, global explanation refers to descriptions of how the overall system works (also referred to as model explanation), and local explanation refers to specific descriptions of why a certain decision was made (outcome explanation). The contributions from this section can be summarized as follows: (1) We present a novel application of Answer Set Programming (ASP) for explaining trained machine learning models. We propose a method to generate explainable rule sets from tree-ensemble models with ASP. More broadly, this work contributes to the growing body of knowledge on integrating symbolic reasoning with machine learning. (2) We present how the rule

set generation problem can be reformulated as an optimization problem, where we leverage existing knowledge on declarative pattern mining with ASP. (3) We show how both global and local explanations can be generated by our approach, while comparative methods tend to focus on either one exclusively. (4) To demonstrate the practical applicability of our approach, we provide both qualitative and quantitative results from evaluations with public datasets, where machine learning methods are used in a realistic setting.

The second contribution of this thesis, presented in Chapter 4, is the development of a differentiable method for computing models of logic programs with non-monotonic semantics in continuous vector spaces. With the recent interest in neuro-symbolic approaches, performing logical inference with linear algebraic methods has been studied as an attractive alternative to symbolic methods. Our interest in the linear algebraic approach to differentiable logic programming is twofold: (1) Modern deep learning approaches rely on real-valued tensor weights internally, so performing logical inference with linear algebraic methods could lead to better integration of symbolic and deep learning methods. (2) Logic programming with continuous values is an active area of research, and there is much to be explored, especially in the context of making logic programming differentiable for neuro-symbolic research.

We build upon previous works by presenting an alternative differentiable method for efficiently computing supported models of normal logic programs in continuous vector spaces. Our main contributions are: (1) Presenting an alternative method for embedding logic programs into matrices and designing an almost everywhere differentiable thresholding function. (2) Introducing a loss function with regularization terms for computing supported models and integrating various gradient update strategies. (3) Demonstrating with a help of systematic performance evaluation on a range of programs, that by selecting appropriate components, it is possible to achieve much higher performance and stability than the existing method.

Finally, the main contributions of this thesis are summarized in Chapter 5. We outline possible paths for future research, including how one might integrate the main results presented in previous chapters, as well as potential challenges.

## 博士論文審査結果

Name in Full  
氏名 竹村 彰浩Title  
論文題目 Knowledge Representation Methods for Bridging Machine Learning and Logical Reasoning

本学位論文は、「機械学習と論理推論の橋渡しのための知識表現手法」（論文和題目）と題し、人工知能(AI)における機械学習と推論の両技術を有機的に結合するために取り組んだ研究内容についてまとめられている。機械学習は近年の深層学習の発展により多方面に応用されているが、説明可能性の問題が指摘されており、知識表現・推論との結合が課題に挙げられている。機械学習と推論を統合することにより、データから記号的知識を機械学習によって効率的にモデル化しロバストな論理的推論が可能となる。本研究ではこの課題に関連して、機械学習モデルを効率よく説明するための論理規則の抽出、及び機械学習との結合を可能とするベクトル空間における論理推論、の二点からアプローチしている。本学位論文は英語で執筆されており全 5 章から構成されている。

第 1 章では本研究の背景として、記号的 AI に関係した知識表現・推論研究と深層学習を中心とする機械学習研究を融合したニューロ・シンボリック AI についての説明があり、それに対する本博士研究での貢献が論じられている。第 2 章では本博士論文で用いられる基礎として、論理プログラミング、解集合プログラミング(answer set programming; ASP)、さらに機械学習結果を説明可能とするための技術について述べられている。

第 3 章は本博士論文の主要成果の一つであり、機械学習で得られるモデルを ASP により説明する手法について述べられている。機械学習モデルとして決定木アンサンブルを取り上げ、各決定木から分類ルールを抽出し、得られた可能なルール集合から ASP の高い表現力を活かして制約・選好に基づいたルール集合を選択する手続きを提案している。決定木アンサンブルから得られたモデルに対しては、全体としてシステムがどのような決定をするかというグローバルな説明、及び特定の決定がなぜ出力されたかというローカルな説明、という 2 種類の説明が得られるように実現している。また ASP の最適化機能を利用することで、少数の重要なルールを生成することを可能としている。提案手法は各種データセットを用いて既存手法と比較し、得られた規則の簡潔性と予測精度の確保という観点から提案手法の有用性を検証している。

第 4 章はもう一つの主要成果として、否定を含む一般論理プログラムにおける標準モデルの一つである支持モデル(supported model)を、微分可能な損失関数の連続最適化に基づいて計算する手法を提案している。深層学習に基づくニューロ・シンボリック AI では記号知識を連続空間上のテンソル表現に埋め込む手法が開発されているが、本手法も同様に論理推論を連続空間上の代数計算により実行するものである。このような微分可能計算を用いて支持モデルを求めることに対し、支持モデル計算に相当する損失関数をうまく設計し、これを最小化することで支持モデルが得られることを理論的に証明している。さらに

論理プログラムの行列表現や連続最適化手法に対して複数の手法を組み合わせた詳細な比較を行っており、結果として既存手法より遥かに計算効率がよく安定した結果が得られることを実験的に示している。

第 5 章では本論文の成果をまとめた後、前章で述べた 2 種の研究課題を統合する方法やその他の今後の課題について述べられている。

公開発表会では博士論文の章立てに従って英語で発表が行われ、その後に行われた論文審査会及び口述試験では、審査員からの質疑に対して適切に回答がなされた。質疑応答後に審査委員会では、出願者は情報学分野の十分な知識と研究能力を持つと認められ、博士研究は機械学習と知識表現・推論の結合に関する研究において十分なレベルの新規性を有しており、学術的に重要な貢献をなしていることが評価された。

以上を要するに本学位論文は、機械学習の出力モデルを説明するための規則を記号推論を用いて生成する技術、及び記号推論を機械学習で用いられるベクトル空間で計算する技術、の 2 点に関して独創的な方式を提案しており、機械学習と記号推論を多面的に橋渡しすることでニューロ・シンボリック AI の新たな可能性を示したものである。また、本学位論文の成果は、2 件の査読付きトップ国際会議フルペーパーとして発表されており、学術的な貢献も認められる。以上の理由により、審査委員会は、本学位論文が学位の授与に値すると判断した。