Learning Symbolic Rules with Neural Guidance

Master’s Thesis Proposal - Fall, Winter and Spring 2025/26

Author: Sarah Lee, MIT / Cal Poly exchange
Advisor: Franz J. Kurfess

Abstract

This thesis explores a hybrid neuro-symbolic framework for rule learning in structured environments. The research combines neural networks’ ability to handle noisy perceptual data with the interpretability and precision of symbolic reasoning. The objective is to enable systems that can extract symbolic rules from neural representations, facilitating explainable AI reasoning. The anticipated contribution is a novel model architecture that aligns deep learning’s flexibility with the rigor of symbolic logic.

Thesis Overview

Modern AI excels at pattern recognition but lacks interpretability and reasoning capabilities. Symbolic AI offers structured reasoning but struggles with perception and adaptability. This research aims to bridge that gap by developing a hybrid system where a neural component guides symbolic rule discovery. The project addresses a significant knowledge gap in explainable learning—how to automatically extract human-understandable symbolic rules from high-dimensional data.

Background

Neuro-symbolic AI has gained traction as a promising paradigm for explainable and generalizable intelligence. Techniques such as Differentiable Inductive Logic Programming (Evans & Grefenstette, 2018) show potential but are limited by scalability. This thesis builds on that foundation, exploring neural guidance to accelerate symbolic rule discovery.

Related Work

Prior work includes DeepProbLog and NeuralLP, which integrate logic reasoning with neural inference. These systems demonstrate feasibility but often require domain-specific tuning. My work extends these models by incorporating self-supervised neural guidance to dynamically adjust rule search priorities.

Contributions

1. A novel neural-guided symbolic rule learner architecture.

2. A synthetic benchmark for explainable rule extraction.
3. Evaluation of interpretability versus performance trade-offs.

Thesis Question / Hypothesis

Hypothesis: Neural guidance can significantly improve the efficiency and accuracy of symbolic rule discovery, resulting in interpretable models without substantial performance loss.

Research Goal and Methodology

The research involves designing a hybrid learning pipeline combining graph-based neural encoders and ILP solvers. Experiments will use logic-based datasets (e.g., CLEVR, bAbI) and measure both reasoning accuracy and rule interpretability. Implementation will be done in PyTorch.

Evaluation and Validation Criteria

Performance will be compared to baseline symbolic and neural methods. Evaluation metrics include accuracy, rule compactness, and explanation fidelity. Human evaluation will assess rule interpretability.

Expected Outcomes and Significance

This work could lead to more explainable AI systems suitable for safety-critical applications like law, science, and robotics. It contributes to the growing field of neuro-symbolic AI by proposing a practical framework balancing reasoning power and interpretability.