universität
innsbruck

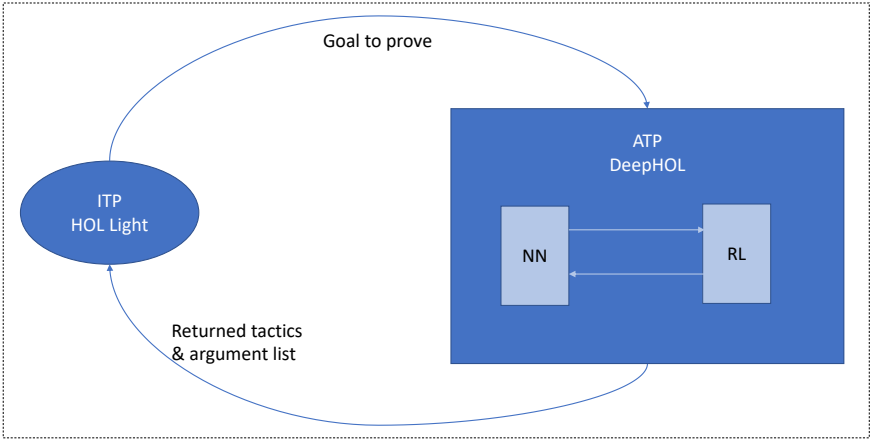
HOList: An Environment for Machine Learning of Higher-Order Theorem Proving

Kamal Zakieldin

Outline

- Introduction
  - Overview
  - Environment
  - Benchmark
- DeepHOL (ATP)
  - Archticture
  - Results
- Conclusion

Overview



HOList Framework

Introduction

- A framework based on the ITP HOL Light.
  - well-defined python API with lots of capabilities.
  - Environment that interact with the ITP HOL Light.
- A benchmark for automated reasoning.
- Automated Theorem Prover (DeepHOL).

Benchmark

- Prepare corpora as a benchmark for theorem proving.
- split dataset into **training** , **validation** and **test** sets of ratio (60%, 20%, 20%).

The three corpora of this benchmark:

	Definitions	Theorems	Proof states
core	240	2320	23512
complex	396	16623	509621
flyspeck	1563	10519	538540
all	2199	29462	1071673

each example in the dataset is a tuple of (goal,tactics,arglist):

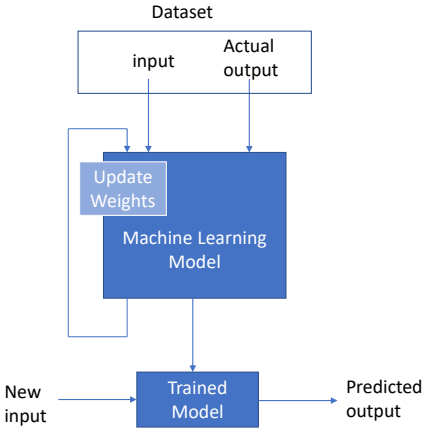
- goal: Theorem to be proven.
- tactics: small set of tactics used to prove that goal.
- arglist: list of theorem that passed as argument to the tactic.

DeepHOL

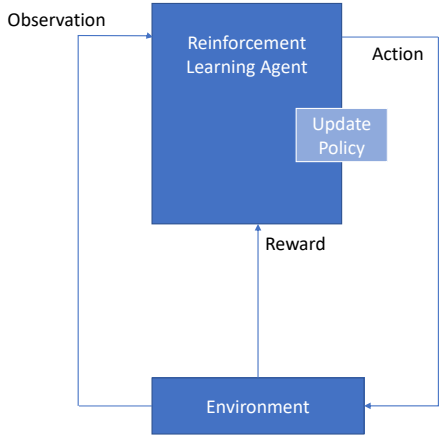
DeepHOL is an Automated Theorem Prover:

- Action Generator.
  - Neural Architecture.
  - Reinforcement Learning Loop.
- Proof Search Graph.

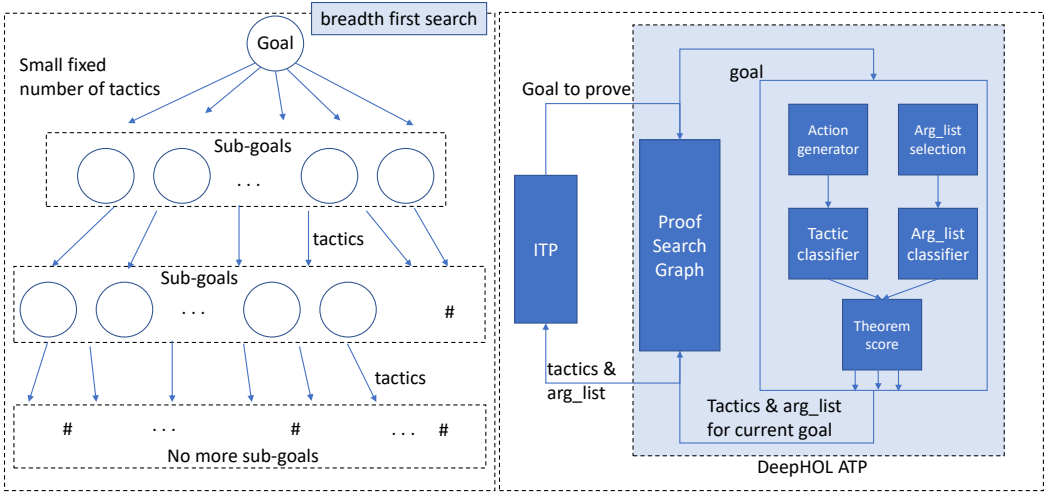
Supervised Machine Learning



Reinforcement Learning



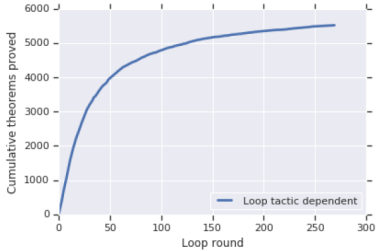
Architecture



results

Description	Proof Success
Built-in Methods	6.1% - 9.2%
NN without Reinforcement learning	17.1% - 24.1%
NN with Reinforcement learning	34.6% - 38.9%

Results in the validation set with 3217 theorems



Results in the training set with 10199 theorems for the NN model with Reinforcement learning

Conclusion

- Automate the formalization of large theories could be possible, as we already know from TacticToe.
- A decent benchmark for automated reasoning.
- Reinforcement Learning are useful and promising.

- **Pros:**
  - Friendly API and an open-source framework.
  - give capabilities to build another archtictures on the framework.
- **cons:**
  - TacticToe still have a better performance and more proved theories.
  - They did not use the TacticToe benchmark to evaluate their work generally.
  - The proofs are not easily readable.

Reference

- K. Bansal, S. M. Loos, M. N. Rabe, C. Szegedy, S. Wilcox, HOList: An Environment for Machine Learning of Higher-Order Theorem Proving.
- https://sites.google.com/view/holist/home.
- Cezary Kaliszyk, Machine Learning for Theorem Proving Course: 703819.
- Justus Piater. Advanced machine learning course: 703642.
- Thibault Gauthier, Cezary Kaliszyk, and Josef Urban. Tactictoe: Learning to reason with hol4 tactics.

Thank you for your attention!

Kamal Zakieldin