Learning Compositional Neural Programs with Recursive Tree Search and Planning

NeurIPS 2019, Spotlight presentation

Thomas Pierrot, InstaDeep

A collaborative research with our partners:

Overview

Our approach

We propose a new Reinforcement Learning agent **AlphaNPI** that uses **modularity**, **hierarchy** and **recursion** as structural biases to **reduce** sample **complexity**, improve **generalization** and increase **interpretability**.

Assumptions

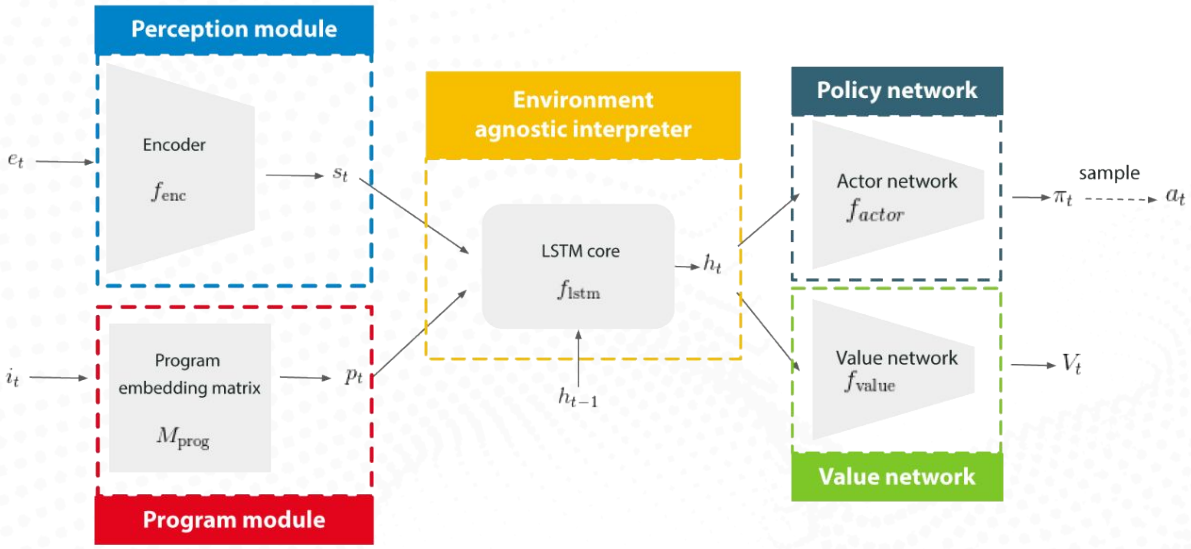
AlphaNPI only assumes a **hierarchical program specification** with **sparse rewards**: 1 when the program execution satisfies the specification, and 0 otherwise.

Experiments

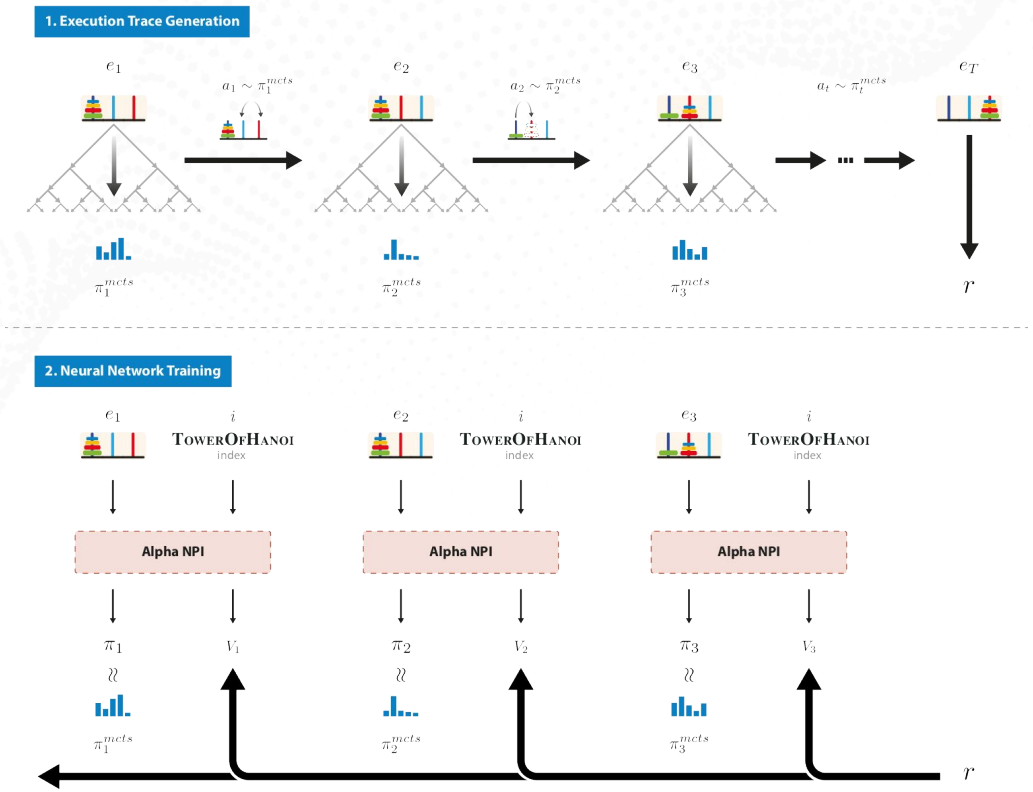
Our agent can **sort large lists** of numbers like its previous NPI version but with lower supervision and **better** sample efficiency. AlphaNPI is also trained to **solve** the **Tower of Hanoi** puzzle with two disks and is proved to **generalize to any number of disks**.

The AlphaNPI algorithm

Our proposed agent, AlphaNPI, **augments the NPI** architecture of Reed and de Freitas [2016] to construct a **recursive compositional** neural network **policy** and a **value** function **estimator**.

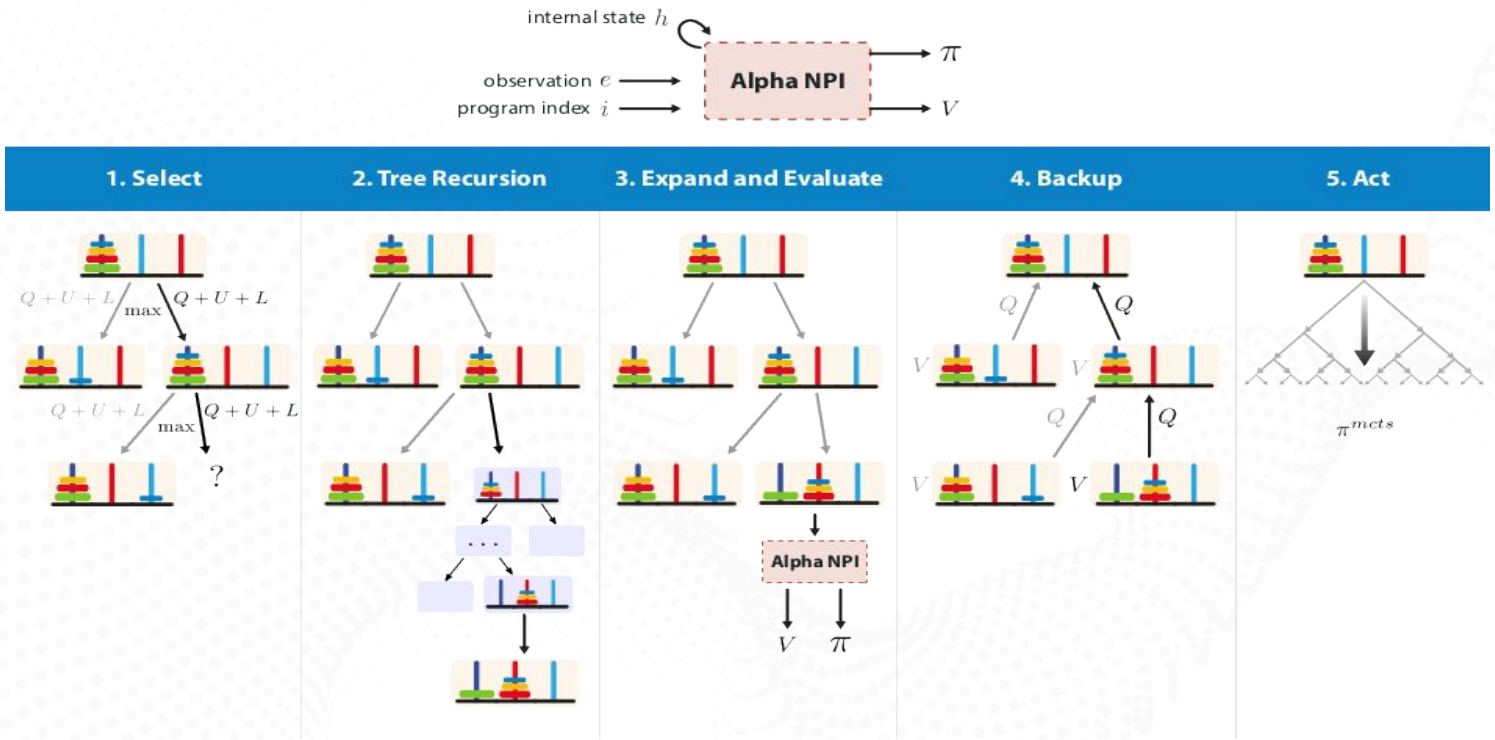


To generate data to **train the AlphaNPI** network by RL, we introduce a **variant of AlphaZero**. The Monte Carlo tree search (MCTS) guided by the AlphaNPI network enables the agent to **"imagine"** likely **future scenarios**.



Adapting AlphaZero to handle compositionality

To deal with **compositionality** we augment AlphaZero with **recursion**.



InstaDeep™

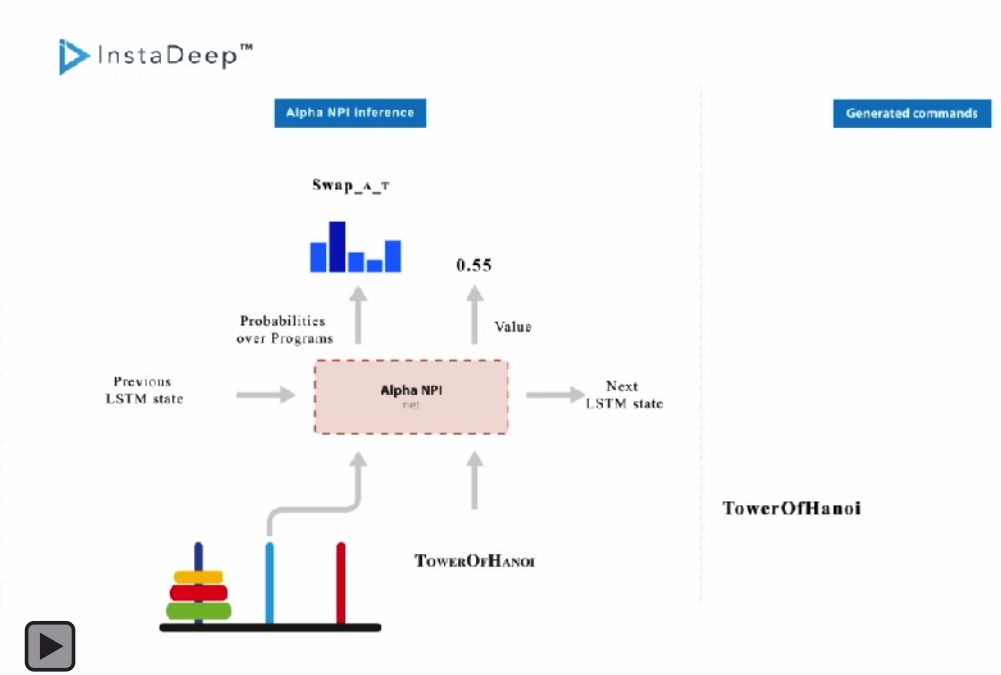
Experiments and results

We trained our algorithm to learn libraries of programs to achieve several tasks:

- Implement the **iterative bubble sort** algorithm **and its recursive version**.
- Solve the **tower of Hanoi** problem **recursively**.

Length	Iterative BUBBLESORT		Recursive BUBBLESORT	
	Net with planning	Net only	Net with planning	Net only
10	100%	85%	100%	70%
20	100%	85%	100%	60%
60	95%	40%	100%	35%
100	40%	10%	100%	10%

Number of disks	MCTS	Network only
2	100%	100%
5	100%	100%
10	100%	100%
12	100%	100%



Thank you!