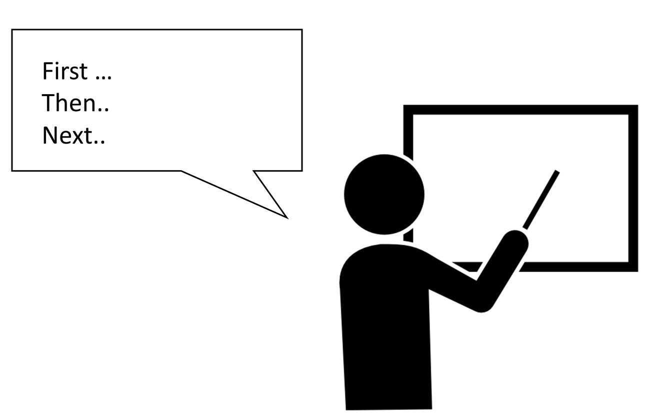
Towards Hierarchical Problem Solving using Natural Language Instructions

Valerie Chen[1] Kenneth Marino[2] Abhinav Gupta[2]

[1]Yale University, [2]Carnegie Mellon University

Motivation

- Sparse reward multi-step problems are difficult for reinforcement learning [4]
- Existing work depends on synthetic languages and templated tasks [1,2,3] and are not generalizable
- However, humans are able to learn from instruction and demonstrations
- Humans are also able to infer similarities and relations from natural language



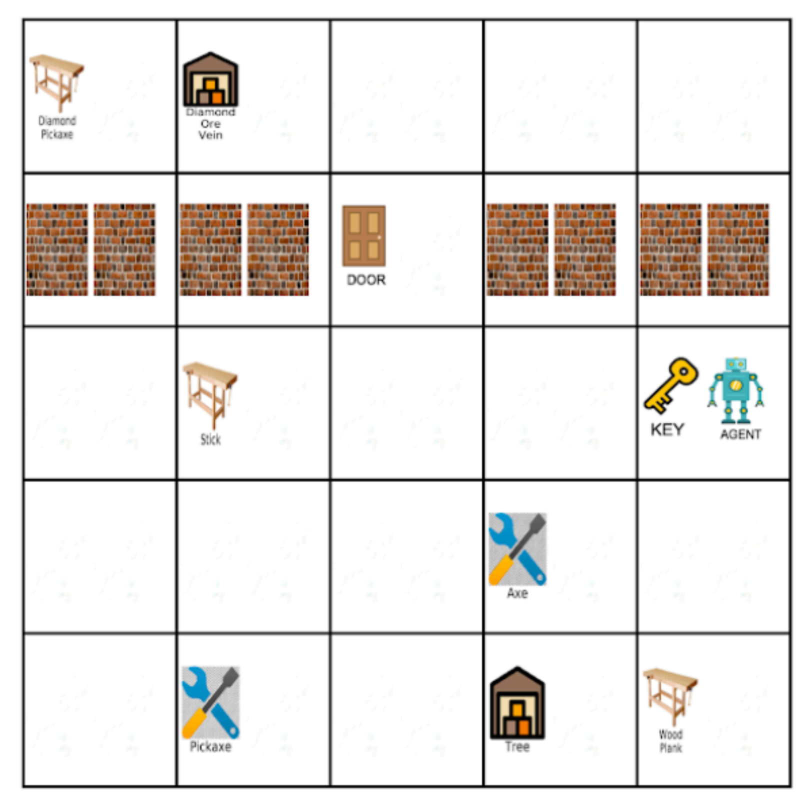
Contributions

- A dataset of annotations from Amazon Mechanical Turk (AMT) of how humans solve complex crafting tasks
- A method that leverages the dataset to guide hierarchical learning algorithms

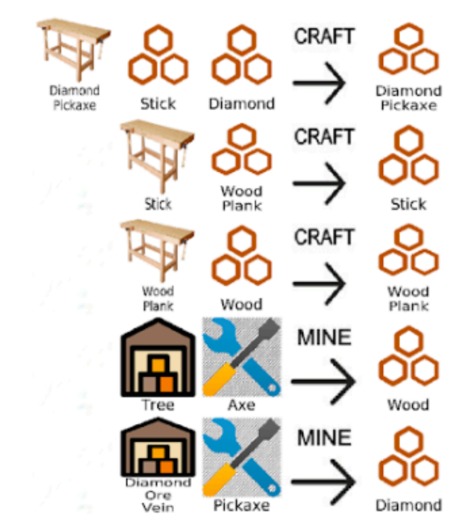
AMT Data Collection Process

Goal
Make a diamond pickaxe

Board



Recipes



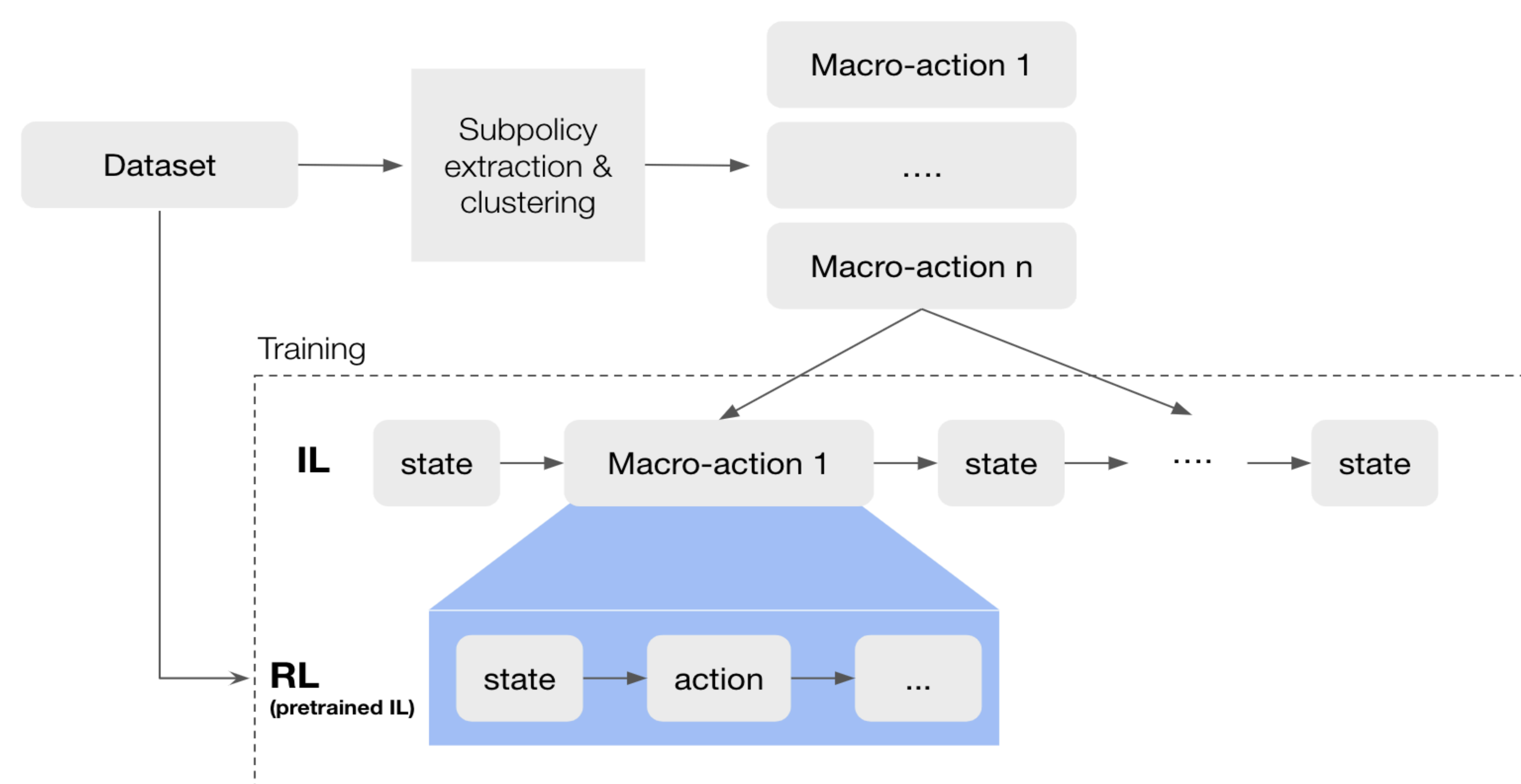
- The worker is given a complex goal, the current board state, and guiding recipes
- The worker must provide step-by-step annotations of how they would achieve the goal accompanied with the appropriate action execution

Data Analysis

- Expected number of traces to be collected: 20,000+
- Total number of unique crafts: 20
- The dataset provides (1) an expert human policy for solving the overall task (2) automatically annotated subpolicies

Proposed Methods

Combining imitation learning and hierarchical reinforcement learning



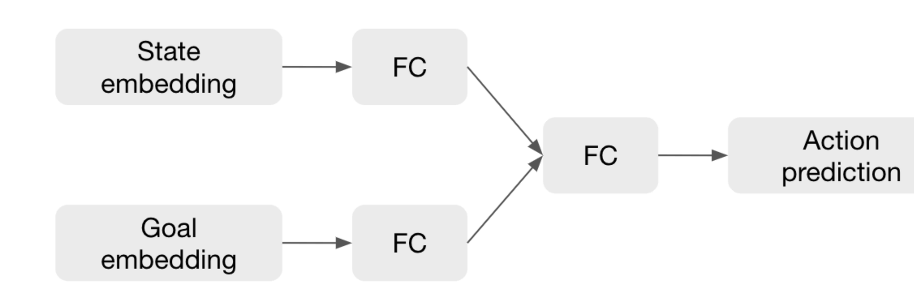
Baseline Comparisons

- Reinforcement Learning: proximal policy optimization (PPO) with sparse reward

$$\theta^* = \arg\max_{\theta} E_{\tau\sim p_\theta(\tau)}[\sum_t r(s_t, a_t)]$$

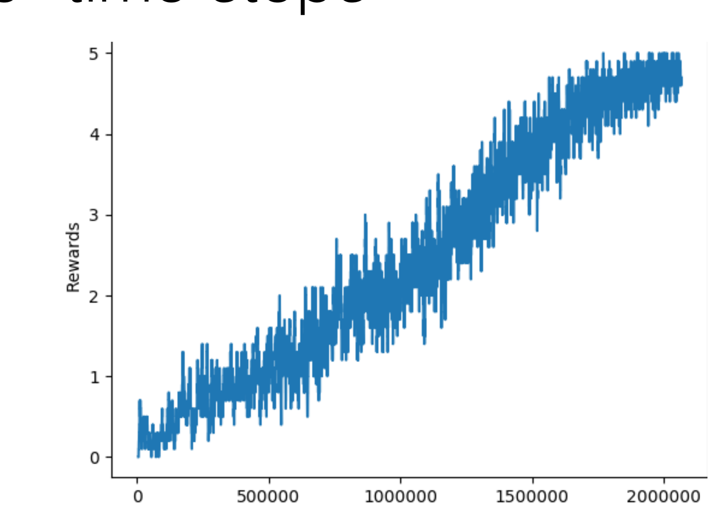
$$r(\tau) = \begin{cases} 1 & goalstate \\ 0 & otherwise \end{cases}$$

- Imitation learning: behavioral cloning with MLP



Baseline Results

- For simple crafting task, PPO takes on the order of 10^6 time steps



Future Work

- Demonstrate generalization to similar yet unseen tasks by using word embeddings
- Improve sample efficiency against baseline methods given only few annotated demonstrations

References

[1] Andreas et al. "Modular multitask reinforcement learning with policy sketches," ICML 2017.
[2] Le et al. "Hierarchical imitation and reinforcement learning," ICML 2018.
[3] Co-Reyes et al "Guiding policies with language via meta-learning," ICLR 2019.
[4] Chevalier-Boisvert et al "BabyAI: First steps towards grounded language learning with a human in the loop," ICLR, 2019.

Acknowledgements

Thank you to Kenny Marino, Abhinav Gupta, Rachel Burcin, John Dolan, and the RISS program for making this collaboration possible.
This work was supported by the National Science Foundation.

Contact Information:
Valerie Chen, v.chen@yale.edu