Towards Hierarchical Problem Solving using Natural Language Instructions

Valerie Chen\textsuperscript{1}  
Kenneth Marino\textsuperscript{2}  
Abhinav Gupta\textsuperscript{2}  
\textsuperscript{1}Yale University, \textsuperscript{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

AMT Data Collection Process

- 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

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

Baseline Comparisons

- Reinforcement Learning: proximal policy optimization (PPO) with sparse reward
  \[ \theta^* = \arg \max_{\theta} \mathbb{E}_{\tau \sim \mu^\theta} \sum_{t} r(s_t, a_t) \]
  \[ r(\tau) = \begin{cases} 1 & \text{goal state} \\ 0 & \text{otherwise} \end{cases} \]
- 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

Proposed Methods

Combining imitation learning and hierarchical reinforcement learning

References


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