Hierarchical RL In StarCraft II with Human Expertise in Subgoals Selection

XINYI XU, TIANCHENG HUANG, PENGFEI WEI, AKSHAY NARAYAN, TZE-YUN LEONG

National University of Singapore

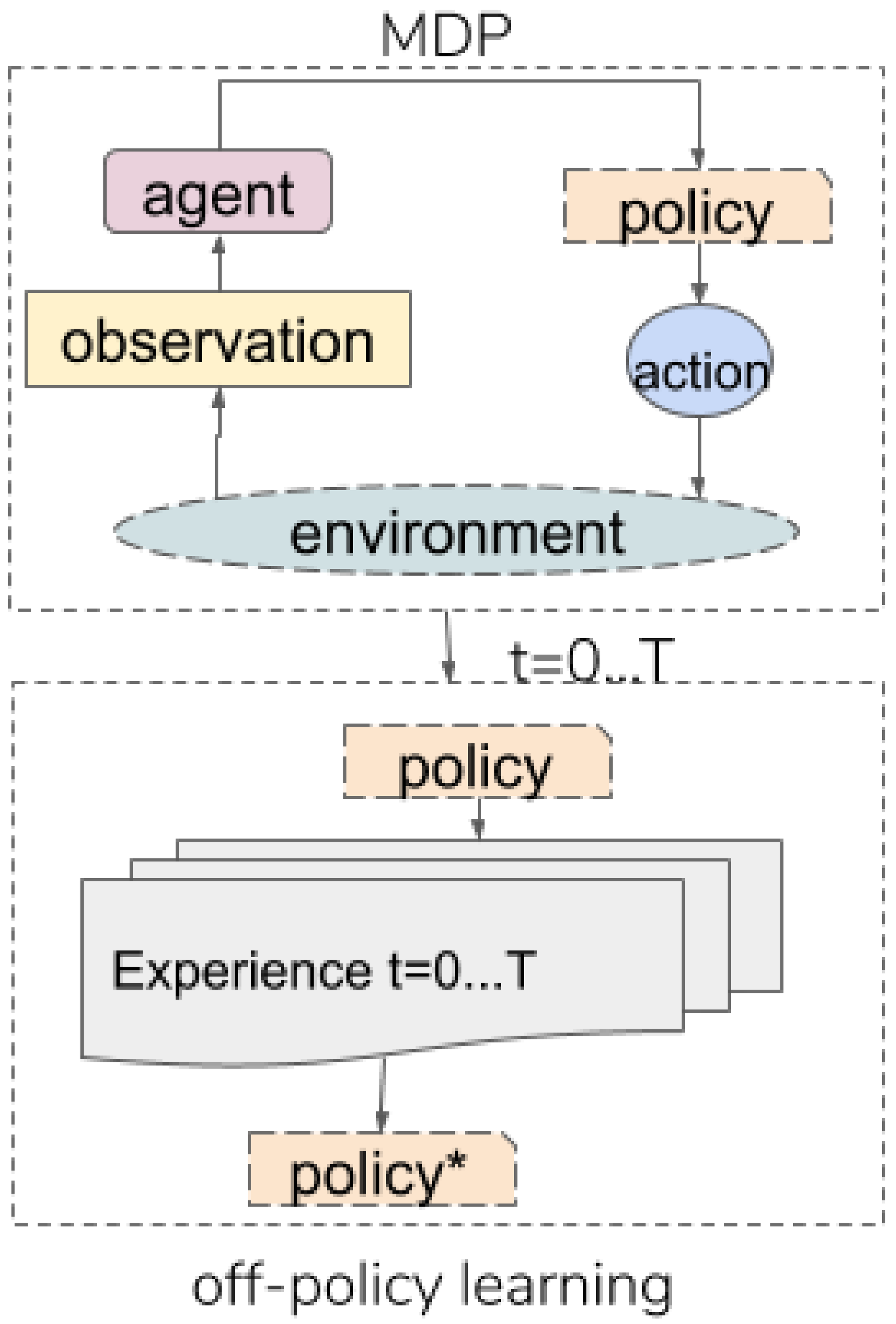
MEDICAL COMPUTING LAB

Introduction

We leverage hierarchical reinforcement learning (HRL) to integrate human expertise in the decomposition of a complex task and implicitly formulate a curriculum. Experimental results in two SC2 minigames demonstrate the sample efficiency and interpretability of our method.

Background

We follow the MDP formulation, use neural networks to represent the value/policy function, and conduct off-policy learning on collected experiences.



Hierarchy

The figure below illustrates the concept of subgoals and subpolicies with a simple navigation agent navigating to the flag post from s_0. Subgoals selected by our method (red nodes) guide the exploration, and contain structural dependence structure (black dashed lines).

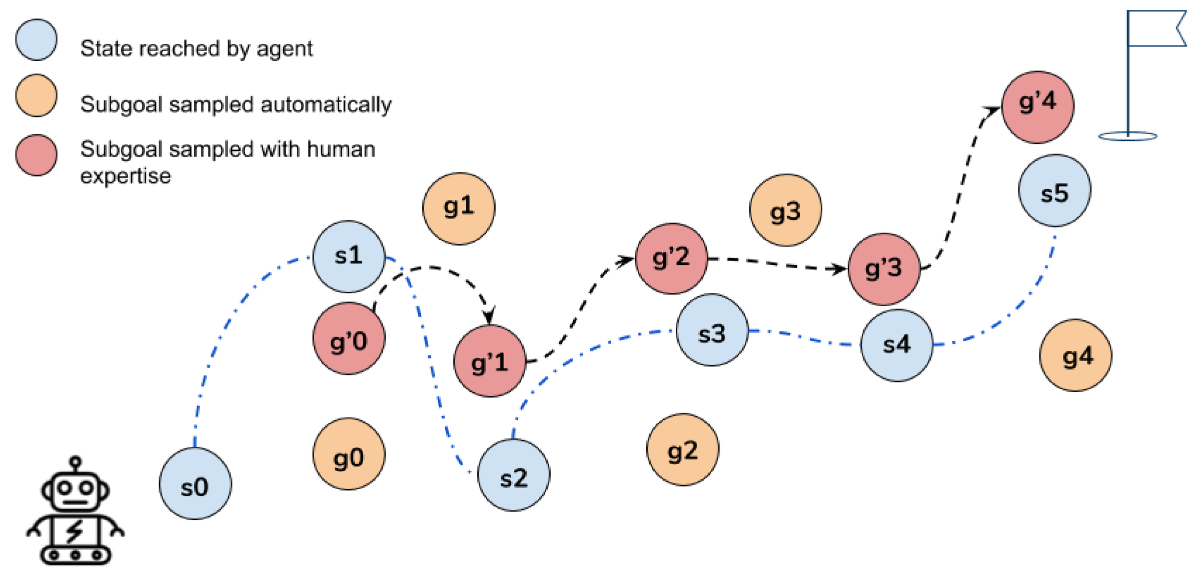


Figure 1: Subgoals and Subtasks

Task Decomposition & Curriculum Design

We implement subtasks by customizing SC2 minigames. For BM we implement 3 subtasks for building supply depots, building barracks, and building marines (with already built barracks), respectively. For CMAG, we have 3 subtasks for collecting minerals, building refineries and collecting gas.

Figure 2: Build Marines. From left to right, top to bottom:(1)-(4): (1) to build supply depots; (2) to build barracks; (3) to build marines with (1) and (2) already built; (4) all three tasks in (1), (2), (3).

Figure 3: Collect Minerals and Gas. From left to right, top to bottom:(1)-(4): (1) to build refineries; (2) to collect gas with built refineries; (3) both tasks in (1) and (2); (4) all three tasks in (1), (2), (3) and collect minerals.

Experiments

Average and max rewards achieved and the number of samples used highlight sample efficiency. The reward curves demonstrate interpretability of the agent's learning and performance.

Minigame	SC2LE	DRL	Ours	Human Expert
CMAG	3,978	5,055	478.5(527)	7,566
BM	3	123	6.7(6.24)	133

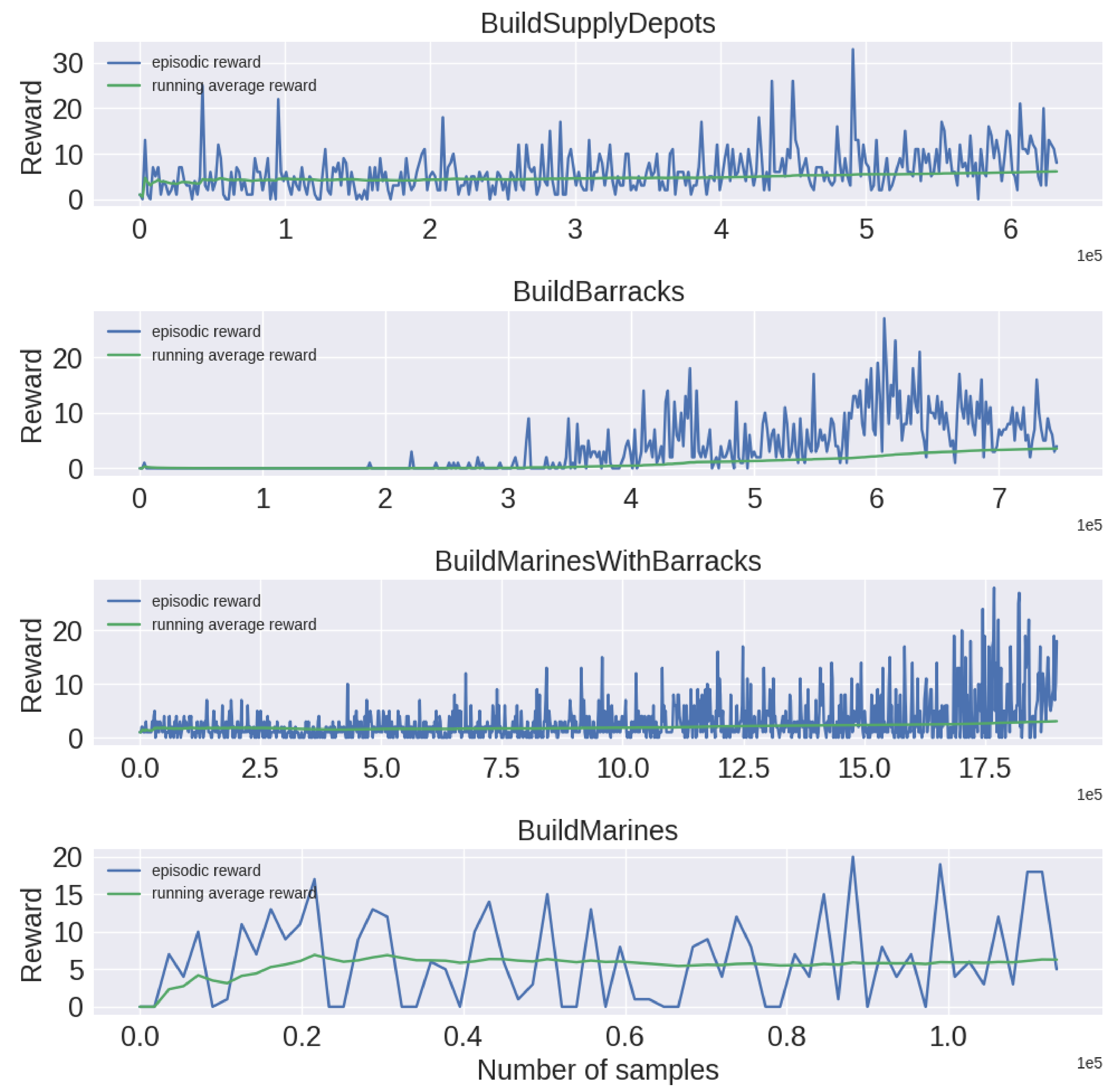
Table 1: Average Rewards Achieved

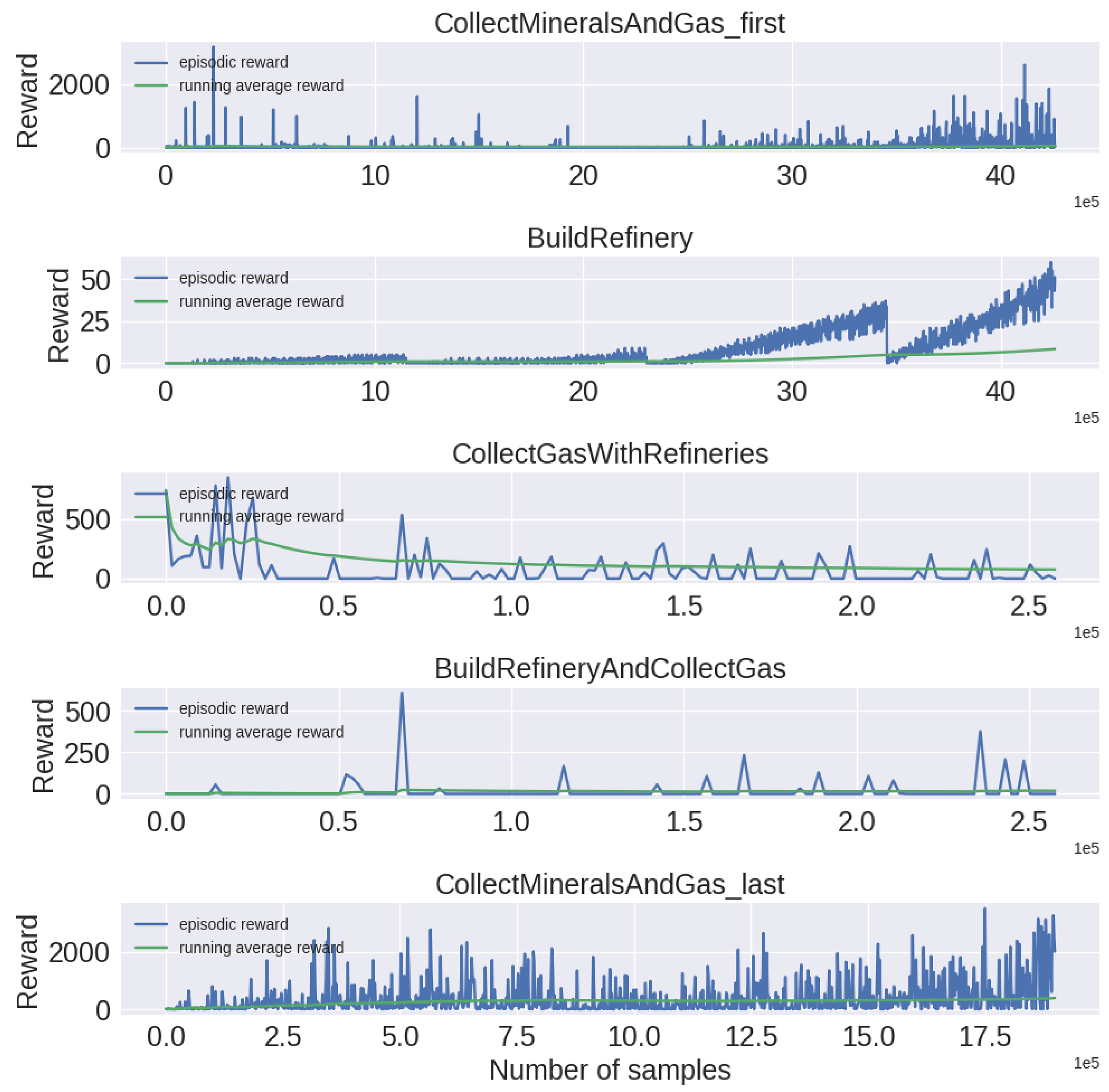
Minigame	SC2LE	DRL	Ours	Human Expert
CMAG	4,130	unreported	1825	7,566
BM	42	unreported	22	133

Table 2: Maximum Reward Achieved

Minigame	SC2LE	DRL	Ours	Human Expert
CMAG	6e8	1e10	1e7	N.A
BM	6e8	1e10	3.4e6	N.A

Table 3: Training Samples Required





Acknowledgement

This work was partially supported by an Academic Research Grant T1 251RES1827 from the Ministry of Education in Singapore and a grant from the Advanced Robotics Center at the National University of Singapore.

Future Research

This initial work invites several exploration directions: developing more efficient and effective ways of introducing human expertise; a more formal and principled state representation to further reduce the complexity of the state space (goal space) with theoretical analysis on its complexity; and a more efficient learning algorithm to pair with the HRL architecture, *Experience Replay* and *Curriculum Learning*.

Contact Information

Web http://medcomp.comp.nus.edu.sg/
Email xinyi.xu@u.nus.edu