High-Throughput Synchronous Deep RL

ILLINOIS

Iou-Jen Liu, Raymond A. Yeh, Alexander G. Schwing

NEURAL INFORMATION PROCESSING SYSTEMS

1. Introduction

Problem:

- Deep reinforcement learning (RL) has **long training time**
- To reduce RL training time, many parallel **synchronous** and **asynchronous** RL methods have been proposed

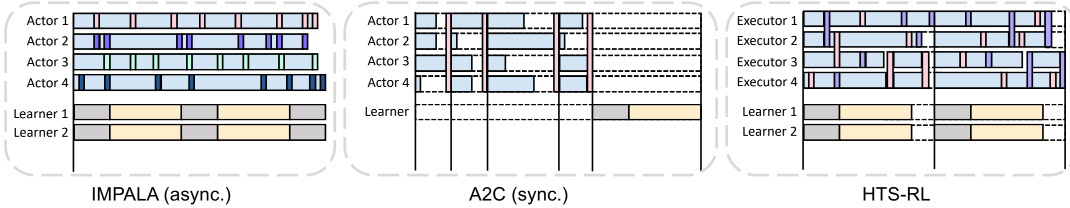
Challenge:

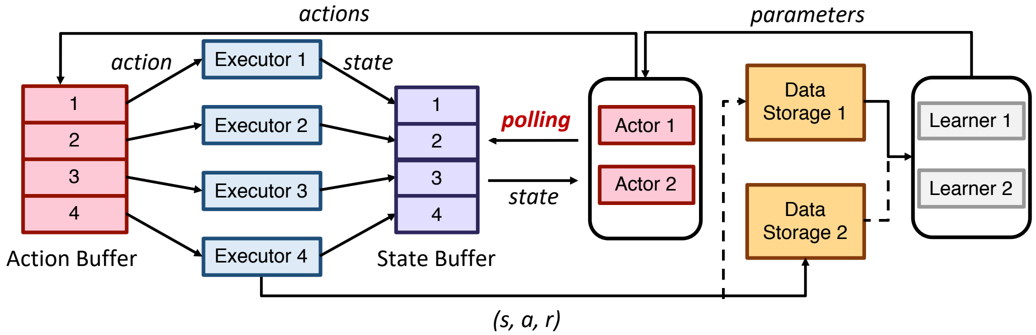
- **Synchronous** methods:
  - Advantage: data efficiency, training stability, full determinism, reproducibility
  - Disadvantage: **synchronization overhead**
- **Asynchronous** methods:
  - Advantage: high throughput
  - Disadvantage: **low data efficiency, non-determinism, stale-policy issue**

Ours: High-Throughput Synchronous RL (HTS-RL)

- Combine advantages of synchronous and asynchronous methods while avoiding their disadvantages
- HTS-RL achieves
  - **High-Throughput**
  - **Without sacrificing data efficiency**

2. Overview





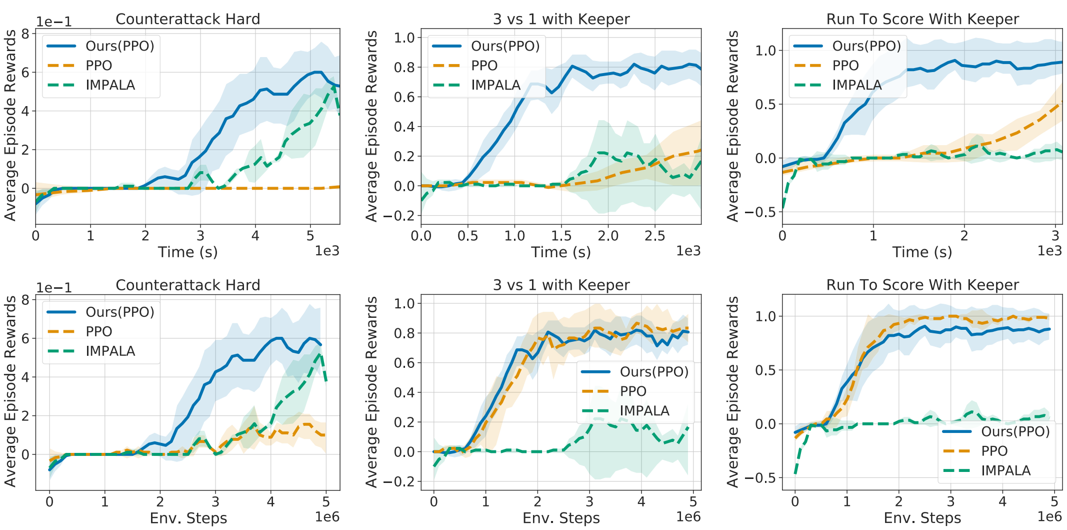
Approach

- Three types of processes: actors, executors, and learners
- Concurrent rollout and learning
- Actors and executors interact in an asynchronous manner, which prevents actors from waiting for executors
- Batch synchronization
  - ➢ Synchronize all processes every alpha steps
  - ➢ Effectively reduces the synchronization overhead
  - ➢ Maintain full determinism

https://ioujenliu.github.io/HTS-RL/

3. Experimental Results (see paper for additional results)

Comparison of HTS-RL, PPO and IMPALA on GFootball environment



Required time (min) to achieve average target scores

- Max score = 1 for each task
- Target score: 0.4 / 0.8

Method	IMPALA	PPO	HTS-RL
Empty goal close	1.7/2.6	5.4/15.5	**1.0/2.0**
Empty goal	8.4/11.7	12.8/19.2	**2.0/3.9**
Run to score	27.0/34.6	16.2/32.5	**6.3/11.4**
RSK	52.3/-	51.2/68.2	**11.5/18.8**
PSK	-/-	70.0/-	**38.8/-**
RPSK	22.3/**25.4**	45.2/90.8	**13.5**/27.1
3 *vs.* 1 w/ keeper	-/-	67.4/144.2	**15.9/25.6**
Corner	-/-	-/-	-/-
Counterattack easy	-/-	223.2/-	**91.3/-**
Counterattack hard	-/-	383.4/-	**61.8/-**
11 *vs.* 11 w/ lazy Opp.	58.2/-	95.8/260.9	**14.4/72.1**