Genetic Algorithm for Curriculum Design in Multi-Agent Reinforcement Learning : Additional Appendix

Anonymous Author(s)
Affiliation
Address
`email`

1 Appendix: Scalability With Respect to Environment Encoding Dimension

1.1 Overview

This section of the appendix examines the scalability of our curriculum generator concerning the dimensionality of environment encoding.

1.2 Benchmark

Given that training a model in a multi-agent environment can take days, we focus on a single-agent environment for this appendix to specifically explore scenario encoding dimensionality.

The selected benchmark is **BipedalWalkerHardcore** [1]. In this environment, an agent uses LiDAR, IMU, and joint encoder data to observe its surroundings and controls the torque on each leg servo of the bipedal walker. Scenarios are defined by the type, size, and location of obstacles (e.g., stairs, pitfalls, and stumps) in the environment. The objective is to navigate through these obstacles without falling. Scenario encoding dimensions range from 20 to 300D, represented by a mixture of integers and floats. Figure 1 visualizes the environment.

Figure 1: Visualization of BipedalWalkerHardcore benchmark

1.3 Baseline

We use Genetic Curriculum (GC) [2] as the baseline for comparison, as GC is one of the top-performing algorithms for the given benchmark.

1.4 Training and Evaluation

We train each algorithm using 5 seeds, with 3.5e7 training steps per seed. Note that the additional time steps used by GC for evaluating and generating the curriculum are not included in the training steps count. We evaluated each trained seed across 1,000 randomly generated scenarios and report the mean performance.

We use the hyperparameters from the GC [2], as the paper is reported based on the BipedalWalker-Hardcore benchmark. Table 1 shows the hyperparameters used.

Start Steps	10000
Learning Rate	3e-4
γ	0.98
α	auto
Batch Size	256
Replay Size	1e6
Curriculum Size	300

Table 1: Selected hyperparameters for the benchmark

Baseline	Mean Reward	Failure Rate
GC	304.33±1.65	**3.96±0.37**
GEMS (Ours)	**311.21±3.56**	**4.27±0.99**

Table 2: Mean reward and failure rate of trained models

1.5 Results

Table 2 shows that while our GEMS does not require additional evaluation steps for curriculum generation, it scales quite well to an environment where the environment encoding can be large as 300D. Table 2 demonstrates that our GEMS, which does not need extra evaluation steps for curriculum generation like GC, effectively scales to environments with encoding dimensions as large as 300D.

References

[1] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

[2] Y. Song and J. Schneider. Robust reinforcement learning via genetic curriculum. In *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022.