Cart2poles: A New Simple Reinforcement Learning Environment Based on IssacGym

Zhai Mingshu [1]

Abstract

IssacGym is a popular framework for reinforcement learning (RL) environments, while Cartpole stands as a classic RL environment. In this paper, the author introduces a novel RL environment built upon the foundation of Cartpole and IssacGym. The environment comprises a 1-DOF slider, a cart, and two interconnected poles. Extensive experiments are conducted on this new environment to evaluate its performance and characteristics. By leveraging the flexibility and capabilities of IssacGym, the author provides a fresh perspective on RL training scenarios, opening avenues for further research and exploration.

1. Introduction

IssacGym(Makoviychuk et al., 2021) is a widely adopted framework for creating and simulating reinforcement learning (RL) environments, while Cartpole stands as a classic RL environment used for benchmarking RL algorithms. However, in this paper, the author introduces a novel RL environment that builds upon the foundation of both Cartpole and IssacGym. This new environment expands the complexity and challenges of the original Cartpole task by introducing additional elements.

The newly developed RL environment consists of a 1-DOF (Degree of Freedom) slider, a cart, and two interconnected poles. The interconnected poles add an extra level of complexity to the task, requiring the agent to learn more intricate control strategies. By incorporating these additional elements, the author aims to create a more realistic and challenging environment for RL training.

To evaluate the performance and characteristics of this novel environment, experiments are conducted. The experiments involve training RL agents using various rewards within the IssacGym framework. The performance metrics, such as convergence speed and stability are thoroughly analyzed and compared with those achieved in the original Cartpole environment.

By leveraging the flexibility and capabilities of IssacGym, the author not only introduces a novel RL environment but also offers a fresh perspective on RL training scenarios. This opens up new avenues for further research and exploration in the field of RL. The enhanced complexity of the environment and the potential for more intricate control strategies provide researchers with an opportunity to investigate advanced RL algorithms, novel reward designs, and innovative training techniques.

Overall, this paper presents a novel RL environment that combines elements from both Cartpole and IssacGym frameworks. The extensive experiments conducted on this environment highlight its performance and characteristics, demonstrating its potential as a challenging and realistic testbed for RL research. The integration of IssacGym provides a valuable platform for exploring advanced RL algorithms and pushing the boundaries of RL training scenarios.

2. Related Work

2.1. Reinforcement Learning

Reinforcement Learning (RL) is a field that trains agents to make sequential decisions by maximizing rewards in an uncertain environment. RL algorithms utilize Markov Decision Processes (MDPs) to model the decision-making problem. Agents learn optimal policies or value functions to guide their actions. Exploration strategies balance between exploring the environment and exploiting learned knowledge. Deep Reinforcement Learning(François-Lavet et al., 2018) (DRL) combines RL with deep neural networks to handle complex, high-dimensional spaces. RL has applications in robotics, game playing, and resource management, among others. Challenges include exploration-exploitation trade-offs and improving sample efficiency. RL research focuses on techniques like curiosity-driven exploration and value function approximation. RL continues to drive innovation in machine learning.

[*]Equal contribution [1]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. Correspondence to: Zhai Mingshu <zhaimingshuzms@sjtu.edu.cn>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

2.2. Issac Gym

IsaacGym, developed by NVIDIA, is a cutting-edge physics simulation environment designed for reinforcement learning (RL) research in robotics and autonomous systems. It offers high-fidelity simulations with realistic physics, enabling the training and evaluation of RL agents for complex robotic tasks. IsaacGym integrates with NVIDIA's deep reinforcement learning framework, Isaac SDK, facilitating the transfer of RL policies from simulation to real-world robots. With its customizable features, efficient performance, and numerous applications in robot manipulation, locomotion, and multi-agent coordination, IsaacGym has emerged as a vital tool in the field of RL and robotics research.

2.3. Cartpole

The Cartpole problem is a well-known benchmark task in reinforcement learning (RL). It involves balancing a pole on a cart, and RL algorithms aim to learn control strategies to keep the pole upright for as long as possible. The Cartpole problem serves as a standard evaluation platform for comparing different RL algorithms and frameworks. It helps researchers explore fundamental RL concepts and assess algorithm performance. By studying the Cartpole problem, researchers gain insights into improving RL algorithms for more complex real-world scenarios.

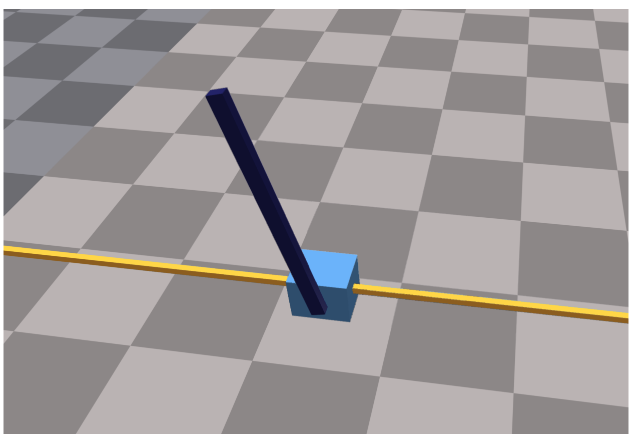

Figure 1. Visualized Cartpole environment in IssacGym

3. Methods

3.1. Build a new environment Cart2pole

To enhance the complexity and challenge of the traditional Cartpole task, an extended version called Cart2Pole has been developed. The Cart2Pole task builds upon the principles of the Cartpole problem by introducing an additional pendulum attached to the existing pole. This modification requires the agent to control both the cart's horizontal movement and maintain balance of the two interconnected poles simultaneously.

In the Cart2Pole task, the agent's objective is to stabilize both poles by applying appropriate control actions to the cart. The agent must carefully manage the movements of the cart to counteract the pendulum's dynamics and prevent both poles from falling. This task demands a higher level of coordination, control, and adaptability compared to the original Cartpole problem.

The Cart2Pole task serves as a more challenging benchmark for evaluating the capabilities of reinforcement learning algorithms. It requires agents to exhibit enhanced decision-making skills, as they need to consider the dynamics and interactions between two connected pendulums. RL algorithms need to learn effective control policies that balance the cart while ensuring stability of both poles, taking into account the increased complexity introduced by the second pendulum.

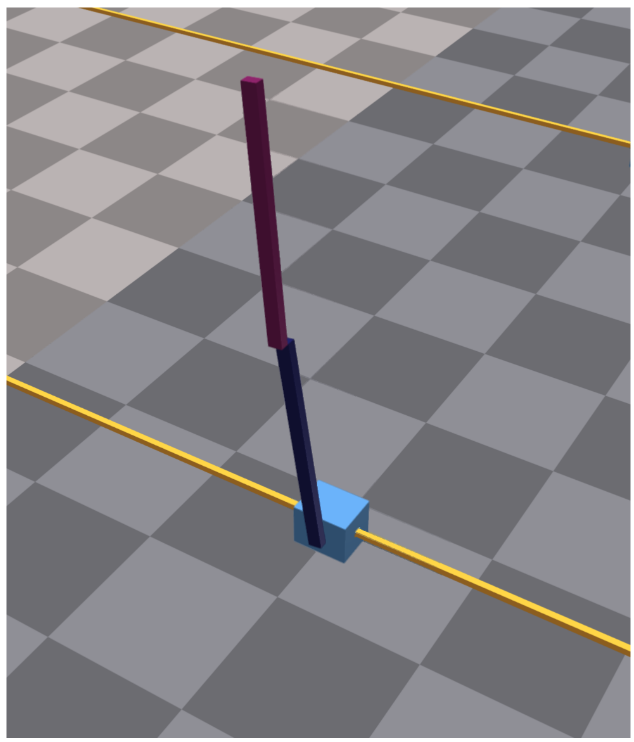

Figure 2. Visualized Cart2pole environment in IssacGym

How to change Cartpole to Cart2pole?

- Build Cart2pole class in code and polish the interface
- Change the XML config file by adding pole and joint.

- Change the training config file to design the training process
- Carefully design a reward to efficiently train Cart2pole

3.2. Reward Design

The design of an appropriate reward function is crucial for the success of reinforcement learning (RL) in the Cart2pole task. In this paper, we propose different reward formulations to effectively train the Cart2pole agent.

3.2.1. Inspired by Cartpole Reward

The Cartpole task utilizes a reward function of the form:

$$r_1 = 1 - 0.01|v_c| - \alpha_p^2 - 0.005|v_p|$$

where v_c represents the cart's velocity, v_p denotes the pole's velocity, and α_p represents the pole's angle. This reward function has been proven effective for training the Cartpole agent.

Building upon this formulation, we naturally extend the reward for the Cart2pole task as follows:

$$r_2 = 1 - 0.01|v_c| - \alpha_p^2 - 0.005|v_p| - \alpha_q^2 - 0.005|v_q|$$

Here, in addition to the Cartpole rewards, we consider the velocity v_q and angle α_q of the additional pole.

3.2.2. Refined Cart2pole Reward

To further refine the reward, we can focus on controlling the behavior of the additional pole to effectively manage the entire Cart2pole system. The final reward used in our experiments is formulated as:

$$r_3 = 1 - \alpha_q^2 - 0.005|v_q|$$

This reward function solely focuses on the angle α_q and velocity v_q of the additional pole.

By refining the reward design, we aim to enhance the learning process of the Cart2pole agent and encourage desirable behavior. The effectiveness of the reward formulations will be evaluated and compared in the experimental section of this paper.

4. Experiment

The experiment utilized the A2C-Continuous algorithm implemented in the rlgames module. The training process was conducted on a single 4GB NVIDIA RTX 3050Ti GPU, with a learning rate of 3e-4 and a total of 250 epochs.

4.1. Comparison between rewards

The results show that r_3 is simpler but more efficient to train Cart2pole robot and the reward curve converges at around 200 epoches.

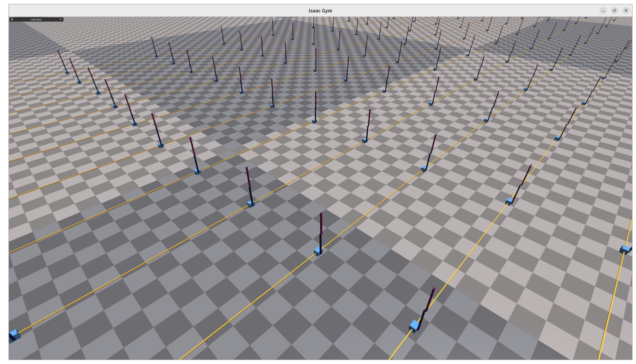

Figure 3. Screenshot during training 512 robots

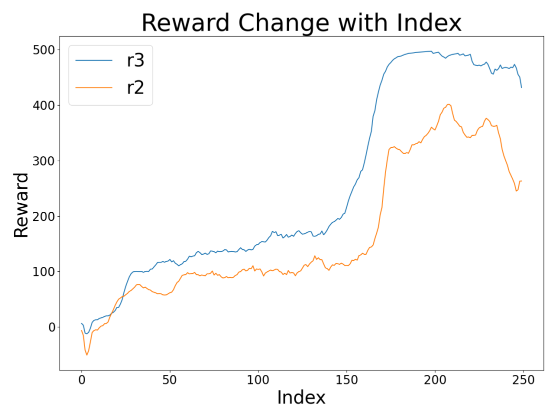

Figure 4. The reward comparison of r_2 and r_3 during training. Note that r_2 is naturally less than r_3 because of more subtractions.But the ideal max reward of both are same 512.Because there are 512 robots training in parallel

4.2. Comparision with Cartpole

Cart2pole simply adds an additional pole to original Cartpole model but it's much more difficult. As you can Imagine, a mankind might know how to control the cart to solve Cartpole but he can hardly solve Cart2pole intuitively.Our experiment with convergence speed support this conclusion.

Accessibility

The code is available at www.github.com/zhaimingshuzms/Cart2pole

References

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., and Pineau, J. An introduction to deep rein-

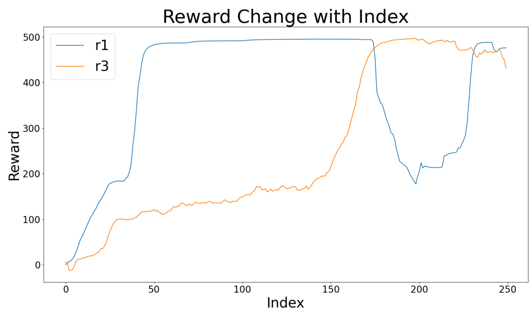

Figure 5. The reward comparison of r_1 and r_3 during training. Cartpole reward converges more quickly than Cart2pole reward. This indicates the difficulty of the environment.

forcement learning. *CoRR*, abs/1811.12560, 2018. URL http://arxiv.org/abs/1811.12560.

Makoviychuk, V., Wawrzyniak, L., Guo, Y., Lu, M., Storey, K., Macklin, M., Hoeller, D., Rudin, N., Allshire, A., Handa, A., and State, G. Isaac gym: High performance gpu-based physics simulation for robot learning. *CoRR*, abs/2108.10470, 2021. URL https://arxiv.org/abs/2108.10470.