Bipedal Walker training and demonstration using reinforcement learning

Pinaq Sharma[1] Nishant P. Singh[2] Mayank Kumar[3] Aditya Bhargava[4]

[1]IIT Jodhpur
m23eet007@iitj.ac.in
[2]IIT Jodhpur
m23irm009@iitj.ac.in
[3]IIT Jodhpur
m23irm007@iitj.ac.in
[4]IIT Jodhpur
m23irm001@iitj.ac.in

Abstract

This report details the implementation of an Evolutionary Strategy (ES) reinforcementlearning algorithm for training a walker agent in the BipedalWalker-v3 environment of OpenAI Gym[2]. The algorithm utilizes a Normalizer class for input preprocessing and a Walker class to generate and evaluate policy variations through positive and negative deltas. The training process involves iteratively refining the policy parameters based on the cumulative rewards obtained from sampled rollouts. The impact of hyperparameters, including learning rate, number of deltas, and noise level, is explored, and periodic video recordings of the agent's behavior provide insights into its learning progress. The report concludes with observations on the algorithm's performance, highlighting both its strengths and potential areas for further refinement in tackling the challenges posed by the complex BipedalWalker-v3 environment.
Keywords: article, template, simple

Contents

1 Introduction

Reinforcement learning (RL) algorithms have demonstrated remarkable success in solving complex control problems, and Evolutionary Strategies (ES) represent a category of approaches that harness the power of evolution to optimize policies. This report presents an in-depth exploration of an ES-based RL algorithm applied to the challenging BipedalWalker-v3 environment provided by OpenAI Gym[2]. The objective is to train a walker agent to navigate the environment effectively by iteratively refining its policy parameters through positive and negative variations. The implementation incorporates a Normalizer class for input preprocessing, and the training process involves generating and evaluating policy variations, selecting promising rollouts, and updating parameters based on cumulative rewards. The report delves into the algorithm's key components, the role of hyperparameters, and the iterative learning process, providing valuable insights into the agent's progress and performance. By addressing the complexities of the

BipedalWalker-v3 environment, this study contributes to the broader understanding of RL algorithms and their potential applications in challenging real-world scenarios.[3]

2 Preliminaries

Before delving into the specifics of the Evolutionary Strategy (ES) reinforcement learning algorithm and its application to the BipedalWalker-v3 environment, it is essential to establish the foundational concepts and context. Reinforcement learning is a paradigm within machine learning where agents learn to make decisions by interacting with an environment to maximize cumulative rewards. Evolutionary Strategies represent a class of optimization algorithms inspired by biological evolution, utilizing random variations and natural selection to iteratively improve solutions. OpenAI Gym serves as the testing ground for RL algorithms, providing a diverse set of environments for benchmarking. The BipedalWalker-v3 environment, in particular, poses unique challenges, requiring agents to learn complex motor control skills for bipedal locomotion. This section sets the stage for a comprehensive understanding of the subsequent discussion by elucidating the fundamental principles of RL, ES algorithms, and the specific characteristics of the BipedalWalker-v3 environment.

3 Methodology

The methodology employed in this study encompasses the implementation and analysis of an Evolutionary Strategy (ES) reinforcement learning algorithm to train a walker agent in the BipedalWalker-v3 environment. The codebase, written in Python, utilizes the OpenAI Gym library[2] for accessing the environment and evaluating the agent's performance. The algorithm is structured around the Walker class, which encapsulates the key components of the ES approach. Input normalization is achieved through the Normalizer class, enhancing the stability and convergence of the learning process.
The training process unfolds in an iterative fashion over a specified number of steps. Random variations, or deltas, are sampled, and the policy's performance is evaluated through positive and negative perturbations. The algorithm then selects the most promising rollouts based on the cumulative rewards obtained. Hyperparameters such as the learning rate, number of deltas, and noise level play a critical role in shaping the learning dynamics, and their impact is systematically analyzed. Video recordings of the agent's behavior are periodically captured to visually assess its progress and proficiency in navigating the BipedalWalker-v3 environment.
The study investigates the sensitivity of the algorithm to hyperparameter choices and provides insights into the challenges faced by the agent during the learning process. The experimental results are presented, showcasing the agent's performance over multiple training iterations. The report concludes with a discussion of the observed outcomes, highlighting the strengths and limitations of the ES algorithm in the context of BipedalWalker-v3 and suggesting potential avenues for future research and improvement.[5][1]

4 Results

The experimental results reveal the evolutionary strategy (ES) reinforcement learning algorithm's dynamic adaptation to the challenges posed by the BipedalWalker-v3 environment. Through a series of training iterations, the agent demonstrates a progressive improvement in its ability to navigate the complex bipedal locomotion tasks. Sensitivity analyses on crucial hyperparameters, including the learning rate, number of deltas, and noise level, provide valuable insights into their influence on the learning process. The agent's performance, as depicted by the cumulative rewards over training steps, reflects the delicate balance required for effective policy optimization. Video recordings at regular intervals offer a qualitative perspective, showcasing the evolving strategies employed by the agent in response to environmental stimuli. The results underscore the algorithm's capability to iteratively refine its policy, offering promising glimpses into its adaptability in the face of a challenging and dynamic task. However, challenges such as convergence speed and occasional suboptimal trajectories also become apparent, prompting consideration for further investigation and fine-tuning to enhance the algorithm's robustness and efficiency.

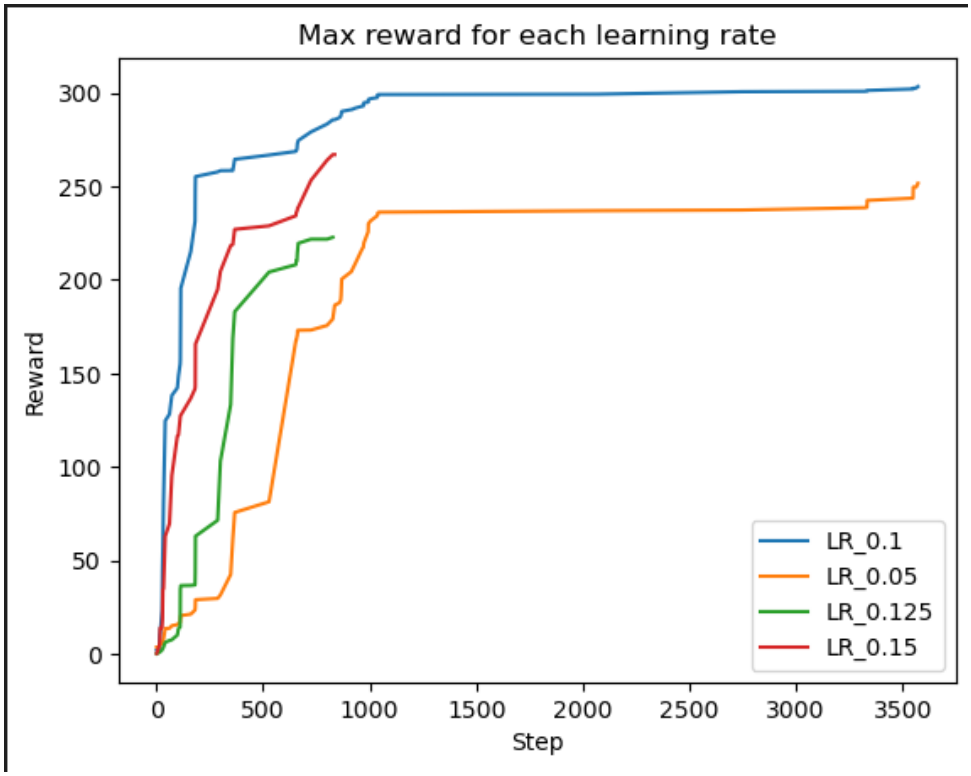

Figure 1: Max reward graph for different learning rate.

5 Conclusion

In conclusion, this study presents a comprehensive exploration of an Evolutionary Strategy (ES) reinforcement learning algorithm applied to the demanding BipedalWalker-v3 environment. The algorithm showcases a notable capacity for training a walker agent to navigate the intricacies of bipedal locomotion. Sensitivity analyses on critical hyperparameters shed light on the delicate interplay between learning rate, number of deltas, and noise level, influencing the algorithm's convergence and effectiveness. The agent's evolving strategies, depicted through cumulative rewards and periodic video recordings, highlight its dynamic adaptation to environmental challenges. While the ES algorithm exhibits promise in policy optimization, challenges such as convergence speed and occasional suboptimal trajectories underscore the complexity of the learning task. These observations suggest avenues for further refinement, potentially through enhanced exploration-exploitation strategies or alternative algorithmic modifications. The insights gained from this study contribute to the broader understanding of ES algorithms in reinforcement learning and provide a foundation for future research aimed at addressing the nuances of complex locomotion tasks in real-world scenarios.[4]

Acknowledgements We would like to express our sincere appreciation to the developers and contributors of the OpenAI Gym library for providing a robust platform that facilitated experimentation in reinforcement learning. The availability of diverse environments has been instrumental in shaping the practical implementation of the Evolutionary Strategy (ES) algorithm in this study. We are also indebted to the authors of the foundational textbook "Reinforcement Learning: An Introduction," Richard S. Sutton and Andrew G. Barto, for their comprehensive insights, which have guided the theoretical framework of this research. Gratitude extends to the broader open-source community for fostering an environment of shared knowledge and collaboration, enriching my understanding of reinforcement learning concepts through online resources and discussions. Special thanks to Dr Anand Mishra for their valuable mentorship, support, and constructive feedback, which have significantly contributed to the quality and depth of this project endeavor.

References

[1] V Akila and J Anita Christaline. Reinforcement learning for walking robot. *Journal Name*, 2017.

[2] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI Gym. 2016. [Online]. Available: `https://gym.openai.com/`.

[3] Jack Dibachi and Jacob Azoulay. Teaching a robot to walk using reinforcement learning. *AA228: Decision Making under Uncertainty*, 2018.

[4] Donghyeon Kim and Glen Berseth. Torque-based deep reinforcement learning for task-and-robot agnostic learning on bipedal robots using sim-to-real transfer. *Journal Name*, 2019.

[5] Jun Morimoto, Gordon Cheng, Christopher Atkeson, and Gerald Zeglin. A simple reinforcement learning algorithm for biped walking. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04*, volume 3, pages 3030–3035 Vol.3, 2004.