Exploring Transfer Learning on Deep Reinforcement Learning Policies

Abstract

This paper investigates whether a simple transfer learning algorithm can be applied to deep reinforcement learning (DRL) policies to decrease training time for new tasks. Specifically, we explore if transferring the initial layers of a pre-trained policy accelerates learning in modified environments. Deep reinforcement learning has achieved significant success in solving complex control tasks. However, training DRL models can be computationally intensive and time-consuming. Transfer learning offers a potential solution by leveraging knowledge from previously learned tasks to accelerate learning in new but related tasks. In this study, we assess whether a simple transfer learning approach can be effectively applied to DRL policies to reduce training time in modified environments.

Introduction

We will test whether we can transfer some knowledge of a DRL policy to be applied to a new task. Using the **Proximal Policy Optimization (PPO)** algorithm on Gymnasium's **BipedalWalker-v3**'s environment, we first train a base DRL model with a neural network architecture of **[24 (input layer) → 64 → 64 → 64 → 64 → 64 → 4 (output layer)]**. We then alter the environment by adjusting **surface friction** and **terrain bumpiness** and retrain a new policy by transferring the first 4 layers from the base policy, freezing their weights and biases, and retraining the last 2 layers. We compare this approach to further training the entire base policy in the new environment without altering any weights or biases initially. We hypothesize that transfer learning will reduce the training time required to adapt to the new environment compared to fine-tuning the entire policy.

Background

Deep Reinforcement Learning

Deep reinforcement learning combines reinforcement learning algorithms with deep neural networks, enabling agents to learn optimal policies in high-dimensional state spaces. **Proximal Policy Optimization (PPO)** is a widely used DRL algorithm known for its stability and efficiency in continuous action spaces.

Transfer Learning in DRL

Transfer learning involves reusing parts of a pre-trained model for a new task. In DRL, this can mean transferring neural network layers from a model trained on one environment to another. Previous research has shown that transfer learning can be effective when the new task shares similarities with the original task.

Hypothesis

We hypothesize that transferring the first four hidden layers of the base DRL policy will reduce the training time required for the agent to achieve comparable performance in the modified environment compared to fine-tuning the entire policy. The transferred layers are expected to provide a strong foundation by retaining fundamental features learned from the original environment, while training new layers allows for adaptation to the new task-specific challenges.

Methodology

Experimental Setup

We trained a base policy using the PPO algorithm on the standard BipedalWalker environment. The policy network architecture consists of:

- Input Layer: 24 neurons (state observations).
- Hidden Layers: Five layers with 64 neurons each, using ReLU activations.
- Output Layer: Four neurons (action outputs).

The base model was trained for 20 million iterations.

Environment Modifications

We created new tasks by modifying the environment:

1. **Slippery Terrain**: Reduced surface friction.
2. **Bumpy Terrain**: Increased terrain irregularities.
3. **Slippery and Bumpy Terrain**: Combined both modifications.

These changes introduced new challenges that the agent needed to learn to navigate.

Transfer Learning Implementation

For the transfer learning approach:

1. **Layer Freezing**: The first four hidden layers of the pre-trained policy network were frozen.
2. **Random Initialization**: The last hidden layer and output layer were reinitialized.
3. **Retraining**: The reinitialized layers were trained on the modified environments, while the frozen layers retained the features learned from the base environment.

Results

Base Model

- **Training Time**: 20 million iterations
- **Mean Reward**: 161.24
- **Mean Time Elapsed**: 1252.48 steps

Transferred Model

- **Slippery Terrain**:
  - **Training Time**: 8 million iterations
  - **Mean Reward**: 158.2
  - **Mean Time Elapsed**: 1228.4 steps
- **Bumpy Terrain**:
  - **Training Time**: 8 million iterations
  - **Mean Reward**: 18.36
  - **Mean Time Elapsed**: 573.4 steps
- **Slippery and Bumpy Terrain**:
  - **Training Time**: 8 million iterations
  - **Mean Reward**: -11.12
  - **Mean Time Elapsed**: 404.97 steps

Discussion

The results indicate that transfer learning can reduce training time in some cases, such as on slippery terrain (a very similar environment), where the transferred policy achieved a reward close to the base policy with less training. However, the performance significantly deteriorated on the bumpy and slippery-bumpy terrains. These results suggest that the transferred layers did not generalize well to environments with more complex dynamics. Overall the results are still inconclusive and we have not yet found an appropriate way to transfer the knowledge.

Challenges

1. **Overfitting**: Transferred layers may overfit to the base environment.

2. **Representation Relevance**: Features learned in the base environment may not be useful in modified environments.
3. **Overparameterization**: The policy network's complexity may hinder effective generalization.

Conclusion

While transfer learning shows promise for deep reinforcement learning, its effectiveness depends heavily on the similarity between tasks. The simple transfer learning approach used in this study failed to generalize to significantly modified environments, highlighting the need for alternative strategies. Future work should explore skill-based transfer learning, where agents learn reusable skills that can be applied to various tasks.

References

Proximal Policy Optimization Algorithms arXiv:1707.06347 [cs.LG]

A Comprehensive Survey on Transfer Learning arXiv:1911.02685 [cs.LG]

Transfer Learning in Deep Reinforcement Learning: A Survey arXiv:2009.07888 [cs.LG]

Skill-based Model-based Reinforcement Learning arXiv:2207.07560 [cs.LG]