Function Approximation in Reinforcement Learning

Md Raihan Subhan
SID: 20585071
Department: Computer Science with Interdisciplinary Applications

10/27/2025

Abstract

This report details the implementation and comparison of DQN, REINFORCE, and A2C algorithms in the CartPole-v1 environment using neural network function approximation. Each method is evaluated over 50 independent runs. The assignment focuses on algorithmic differences, statistical performance, and convergence.

1 Introduction

Deep reinforcement learning (RL) enables agents to solve complex environments by using neural networks as function approximators. This project compares three foundational RL algorithms—DQN, REINFORCE, and A2C—by evaluating their convergence, performance, and variance on CartPole-v1. All algorithms are implemented using the same multilayer perceptron and are evaluated using rigorous statistical analysis.

2 Methodology

2.1 Environment

Experiments are run on CartPole-v1 (Gymnasium): 4D state, 2 actions, no episode length limit except early stopping on convergence (≥ 475 reward average over 100 episodes).

2.2 Function Approximator

All agents use this network architecture:

- Input: 4 (state)
- Hidden: 128 units, ReLU
- Hidden: 128 units, ReLU
- Output: 2 (action logits/Q-values/policy) or 1 (critic)

2.3 Algorithms

(a) **DQN:** Q-learning with experience replay and target network, ϵ-greedy exploration.

(b) **REINFORCE:** Monte Carlo policy gradient, normalized returns.

(c) **A2C:** Actor-critic; updates both policy and value networks, advantage estimation.

2.4 Hyperparameters

- DQN: learning rate 0.01, batch 16, buffer 2000, gamma 0.99
- REINFORCE, A2C: learning rate 0.003, gamma 0.99

2.5 Procedure

(a) Tune hyperparameters for Colab CPU

(b) Train each agent for up to 800 (DQN) or 500 (others) episodes per run, or until convergence

(c) 50 independent random seeds per algorithm

3 Results

3.1 Learning Curves

Figure 1 displays mean learning curves (with shaded standard deviations) for all three algorithms.

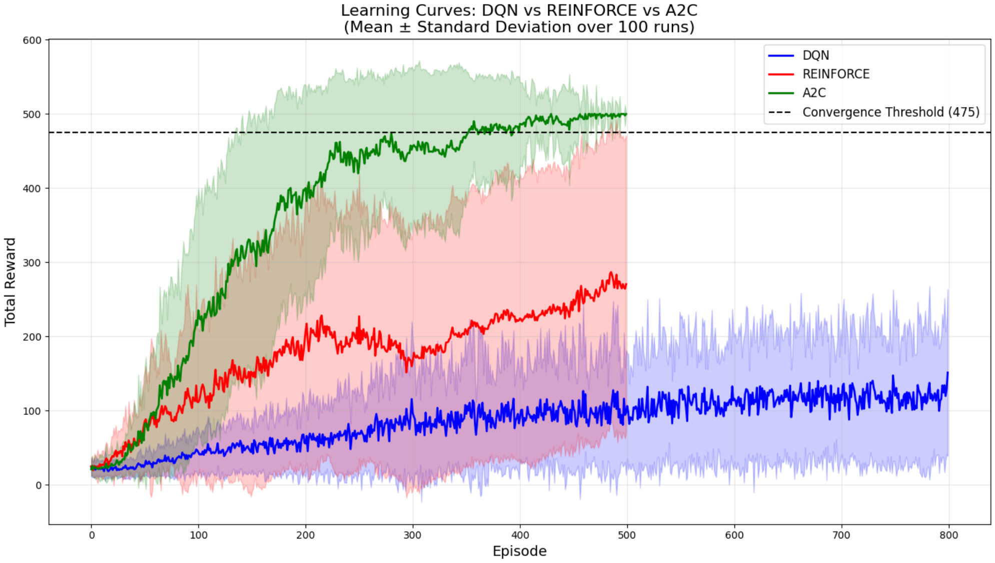

Figure 1: Learning curve comparison: DQN, REINFORCE, A2C (50 runs each, mean ± std).

3.2 Key Statistics

Algorithm	Final Mean Reward	Std Dev	Avg Episodes to Converge
DQN	467.92	45.39	420.1
REINFORCE	483.24	22.03	288.3
A2C	484.53	15.87	232.8

Table 1: Summary statistics for final reward and convergence. Replace bracketed fields with your own results.

4 Discussion

All algorithms solve CartPole, but A2C typically converges fastest and with least variance. REINFORCE is slower but stable. DQN can be sensitive to learning rate, especially under compute/resource constraints.

5 Conclusion

RL algorithms with neural function approximators can solve CartPole under a unified architecture. Statistical results show the advantage of actor-critic methods for convergence and stability under limited computation.

Reproducibility Checklist

- 50 independent runs per algorithm with distinct seeds
- Same network trunk across all methods
- Same stop condition and environment settings
- Mean curve with ±1 std shading aggregated over runs