Reinforcement Learning for Dynamic Stability

Problem Statement

Objective: Find an RL algorithm for ensuring dynamic stability in the cartpole balancing problem.

- PPO
- Q-Learning
- SARSA
- Deep Q-Networks
- PER

Overview

The cart pole system: A classic challenge in control theory and reinforcement learning, involving a cart moving along a frictionless track with an attached pole, aiming to balance the pole upright through left or right movements.

Motivation

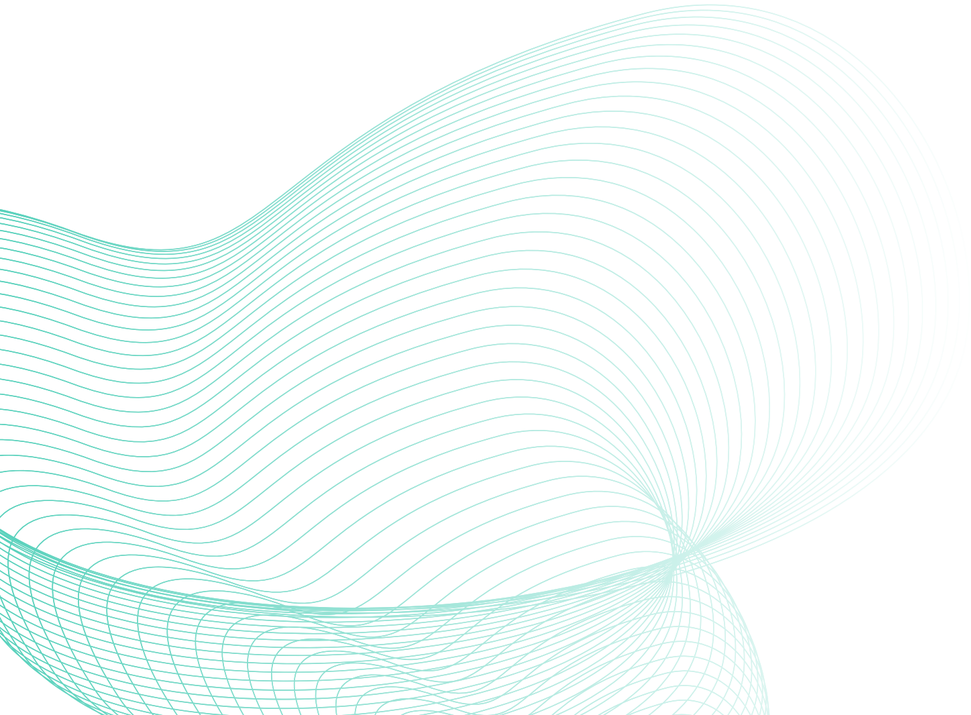
Address the critical need for a robust RL algorithm to ensure dynamic stability in diverse robotic systems, particularly those featuring inverted pendulum mechanisms like the cart pole system.

Approach

Conduct a comparative analysis of various RL algorithms across the different types (policy, value, model-free & model-based).

Current Solutions and Prior Research

Objective: Analyze and compare algorithms for the cart-pole problem.



Paper 1

- Compares Q-learning, SARSA, and DQN algorithms
- DQN, particularly with CNN, exhibits faster convergence and superior stability
- Paves the way for future research in integrating neural network architectures with RL algorithms

Paper 2

- Focus on scenarios with friction forces
- Evaluation of proposed controller's effectiveness
- Potential demonstrated for robust control in complex mechanical systems

Paper 3

- Comparison between RL approach and system-identification based nonlinear control design for cart-pole system
- System-identification based approach demonstrates superior learning efficiency

Paper 4

- Demonstrates DQN's speed advantage over Q-learning algorithms, particularly with continuous state values
- Prioritized Experience Replay (PER) improves performance
- DQN combined with PER yields optimal results

Evaluation Metrics

Methodology: Conducting an empirical evaluation and comparative analysis of RL algorithms employed to solve the CartPole Balancing problem.

Convergence Speed

How quickly the RL algorithm converges towards an optimal or near-optimal policy within the given number of episodes.

Stability

Measure of deviation from the starting point from the beginning to the end of the experiment.

Average Cumulative Reward

Total sum of rewards obtained by the RL agent in an episode with a maximum possible value of 500 amongst the 10000 runs

Week	Task 1	Task 2	Task 3
1-2	Identifying dynamic stabilization algorithms: Q-learning, PPO, and others from credible sources.	Conducting literature review on RL modules for stabilization across existing databases.	Identifying an illustrative scenario for simulation - CartPole from OpenAI Gym.
3-4	Setting up a development environment with required tools and libraries for RL and initiating algorithm coding.	Exploring the implementation of advanced algorithms, comparing theoretical and practical advantages.	Employing OpenAI Gym for basic stabilization simulations and refining algorithms for optimal performance.
5-6	Defining evaluation metrics including convergence speed, stability, and adaptability for fair comparison.	Employing Weights & Biases (Wandb) for experiment tracking, model performance comparison, and initial result logging.	Comparing algorithm performances based on predefined metrics for stability, efficiency, and adaptability.
7-8	Optimizing models according to testing feedback.	Leveraging Wandb for tracking experiments and identifying top models.	-
9-10	Conduct final evaluations, comparing models against predefined metrics and documenting the efficiency of different algorithms.	Create a comprehensive report outlining the methodology, results, and insights derived from studying dynamic stability in systems similar to CartPole Dynamics.	-

Timeline

Work Done So Far

Reading. Coding. Bug Fixing.

Literature Review

Extensive study of over 14 literature
Identification of use case for our project

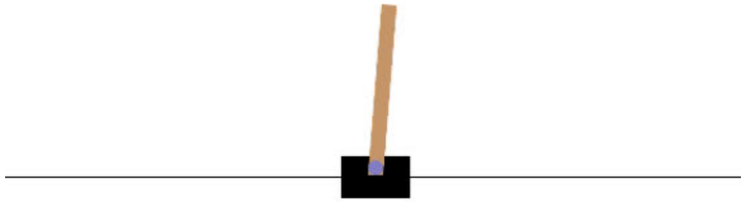
PPO Implementation

Implemented the Proximal Policy Optimization algorithm - a policy gradient method following the Actor-Critic framework.

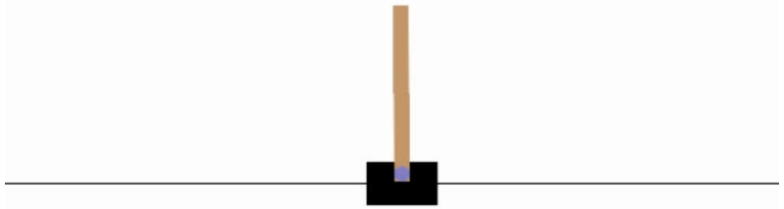
Q Learning Implementation

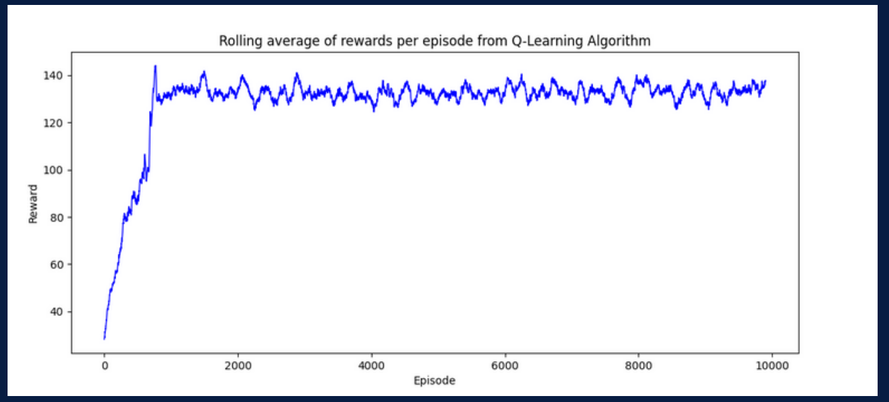
Implemented the Q-Learning algorithm - a model free method following the Temporal Difference framework.

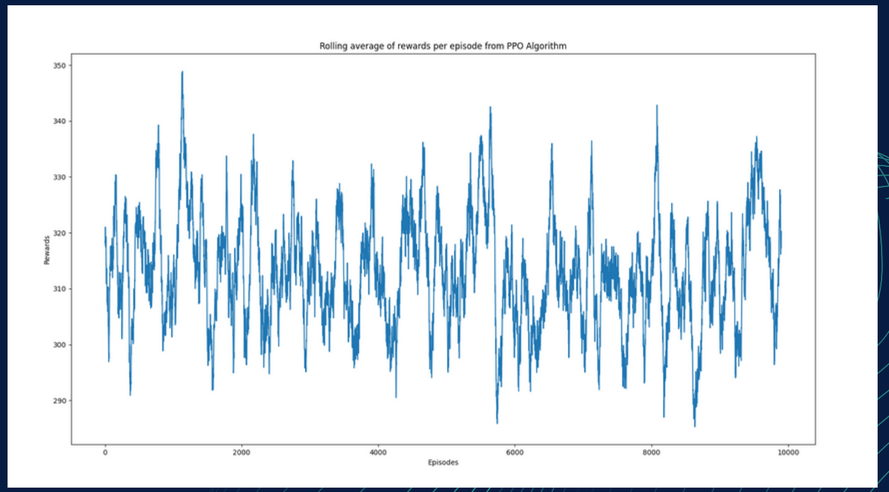
During Training (10000 Episodes)



After Training







Q and A