Neuromorphic Solutions to Reinforcement Learning problems

Sri Ranganathan Palaniappan, Tejeshwar Natarajan, Rahul Narayanan, Chacko Abraham, Jeongyeop Han, Mugdha Daftardar, Dorsa Ajami

Introduction

The Neuromorphic Team at Rogues Gallery specializes in neuromorphic, also known as brain-inspired, applications. We're exploring Q-learning and Deep Q-learning models for the Mountain car environment - a classic problem in reinforcement learning provided by the OpenAI gymnasium library to demonstrate the effectiveness of our models. Q-learning provides a simple, yet effective method for the Mountain Car to leverage acceleration and learn the best sequence of actions to reach the goal.

Methodology

Q-Learning

- Method of reinforcement learning that learns to choose the best action in different states by observing the consequences of its actions and estimating the future reward for each potential action.
- Observation space is divided evenly into discrete chunks for the Q-Table, with each chunk representing a cell, and each cell holding different q values for each possible action.
- Values in the Q-Table are randomly initialized, and for each action taken, values begin to update based on the Q-Learning equation.
- Q-Learning equation determines the action that results in the highest reward for the agent at a given position, and updates the entry for it in the Q-Table with a higher value.

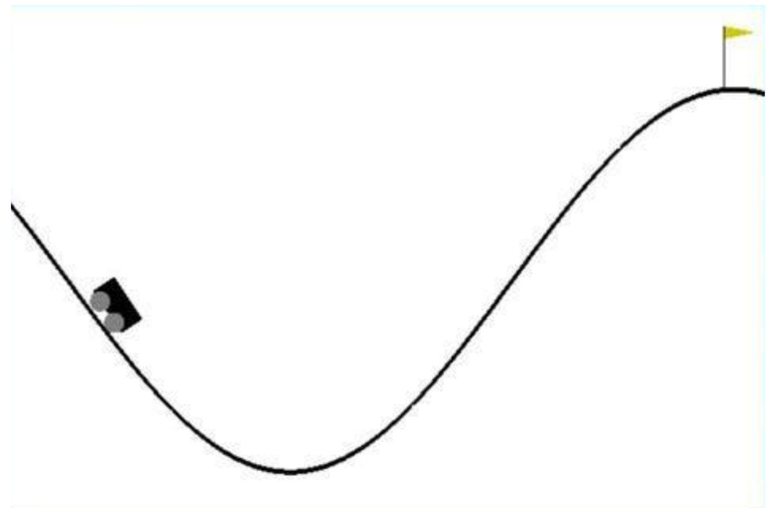
$$Q(s_t,a_t) \leftarrow (1-\alpha)\cdot \underbrace{Q(s_t,a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1},a)}_{\text{estimate of optimal future value}}\right)}^{\text{learned value}}$$

Q-learning equation

Key Experiments & Results

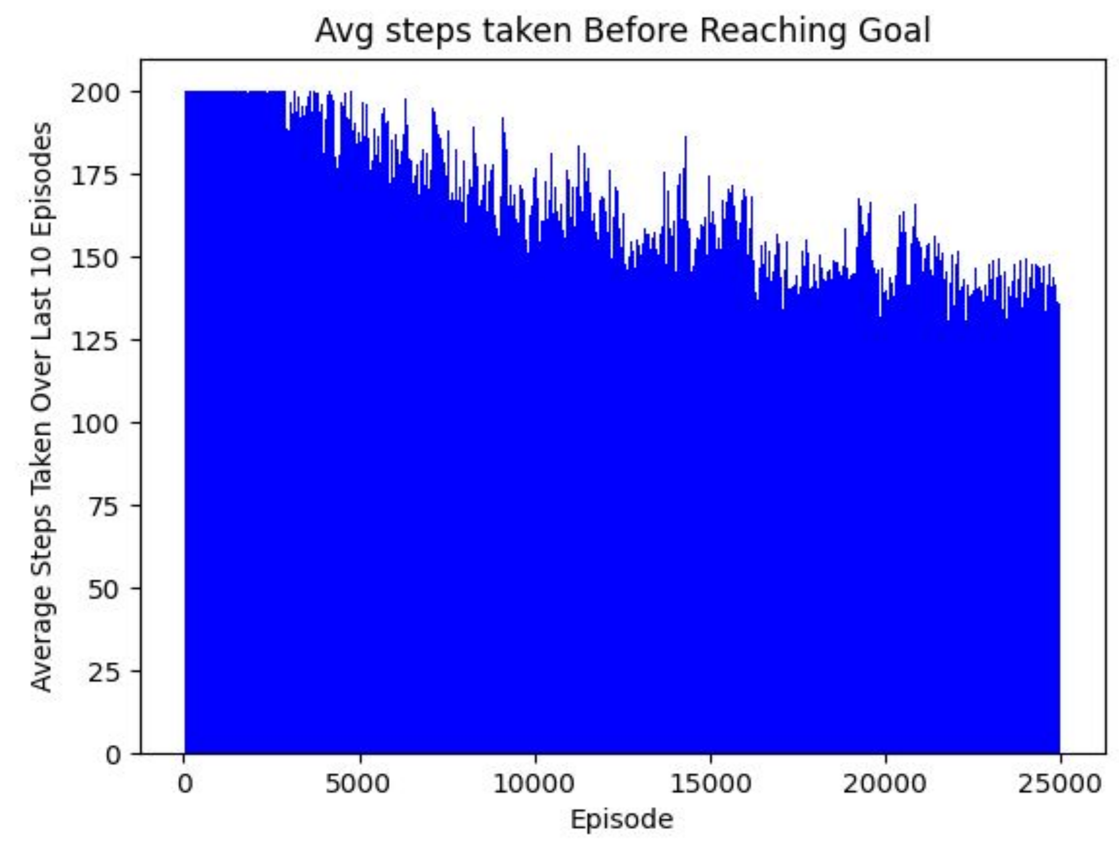
Mountain Car Problem

- The mountain car environment requires a car (agent) with insufficient power to ascend the steep hill to reach the flag.
- The only available actions for the car are accelerations on either directions.
- The car must learn optimal actions to strategically enough momentum and reach the flag placed on top of the right hill.



Hyperparameter Optimization

- Employed grid search over a valid range for learning rate and discount factor to get all combination of values and evaluate the agent's performance over a series of episodes.
- Evaluated average time to reach the goal over a series of episodes as the "score" (primary performance metric).
- Fine-tuned parameters resulted in improved stability and faster convergence of the Q-learning algorithm.



Challenges & Next steps

Challenges

- Encountered difficulties in rendering the environment due to reliance on outdated documentation for the OpenAI Gym framework instead of the updated Gymnasium version.
- Experienced challenges in determining the optimal epsilon value during the hyperparameter tuning phase.

Next steps: Deep Q-Learning

- Conducted an in-depth investigation into the distinctions between Q-learning and Deep Q-learning, focusing on understanding the nuances involved in learning optimal policies.
- Develop a DQL algorithm implementation for the Mountain Car problem, which can handle continuous state spaces, employing a neural network as a Q-function approximator, implementing experience replay, and incorporating a target network.