RL Research Report

Abstract

This report presents a Q-learning-based reinforcement learning solution applied to the FrozenLake-v1 environment. Following the DOT framework, the report outlines the design motivation, implemented operations, and tested outcomes through experimentation. Key focus was placed on hyperparameters, exploration strategies, and the goal reach rate to evaluate agent learning performance. The experiments were documented and interpreted using code linked in the accompanying repository.

Design

The main objective of this project was to apply Q-learning in a discrete reinforcement learning environment to gain a practical understanding of agent learning. The FrozenLake-v1 environment was selected due to its simplicity, yet non-trivial dynamics. This assignment aimed to explore how exploration strategies and learning parameters affect the learning curve, policy quality, and goal reach rate of the agent.

Operation

The agent was implemented using the Q-learning algorithm in Python, applied to the FrozenLake-v1 environment from the Gymnasium library. Two exploration strategies were tested: epsilon-greedy and Boltzmann softmax. The experiments were carried out by adjusting key hyperparameters such as learning rate (alpha), temperature (for Boltzmann), and enabling/disabling environment slipperiness. The implementation also tracked the agent's goal success rate (y variant), and visualizations were produced for reward curves, policy heatmaps, and exploration comparisons.

Test

Several experiments were conducted to interpret the effects of different variables on learning. The agent struggled more in the slippery environment, and learning was slower due to stochastic transitions. Epsilon-greedy exploration showed more stability than Boltzmann unless the temperature was fine-tuned. Higher learning rates accelerated early learning but introduced instability. The goal reach rate remained low, explaining the weak policy in the final Q-table visualization. These experiments highlight the importance of environment structure, exploration tuning, and reward signal frequency in reinforcement learning.