Report of the 3rd Reinforcement Learning practical

Pierre-Louis Favreau - Samy Amine

Introduction

In this project, we implemented reinforcement learning agents to solve the OpenAI Gym's Taxi-v3 environment, a grid-based game where the agent aims to pick up and drop off a passenger at specific locations while minimizing actions taken. The implemented algorithms include Q-Learning, SARSA, and an extension of Q-Learning with epsilon scheduling to analyze their learning efficiencies and performances.

Methods

We utilized three algorithms for the agent: Q-Learning, SARSA, and Q-Learning with epsilon scheduling. The environment was run for 1000 episodes for each algorithm. All agents were developed with the following hyper-parameters: learning rate (α) of 0.1 or 0.5, discount factor (γ) of 0.99, and an epsilon-greedy policy with varying epsilon values for exploration.

The Q-Learning agent utilized a greedy approach based on learned Q-values, while SARSA (State-Action-Reward-State-Action) updated the Q-values by considering the next action chosen by the agent in each update. Additionally, a Q-Learning agent with epsilon scheduling was implemented to gradually reduce exploration over time, improving exploitation after sufficient learning.

Results

The learning curves for all three algorithms are shown in Figure 1. The convergence patterns of Q-Learning, SARSA, and Q-Learning with epsilon scheduling are largely similar, with each algorithm converging to comparable average rewards. Q-Learning demonstrated a slightly faster early convergence due to its off-policy nature, which promotes aggressive updates. However, overall, the final rewards achieved by all agents were very similar, highlighting that each of these methods is capable of successfully learning the optimal policy for the Taxi-v3 environment.

Conclusion

All three algorithms—Q-Learning, SARSA, and Q-Learning with epsilon scheduling—successfully converged to optimal policies for the Taxi-v3 task, with comparable final performance. Minor differences in early convergence speed were observed, but ultimately, all algorithms demonstrated similar effectiveness in learning the optimal policy.

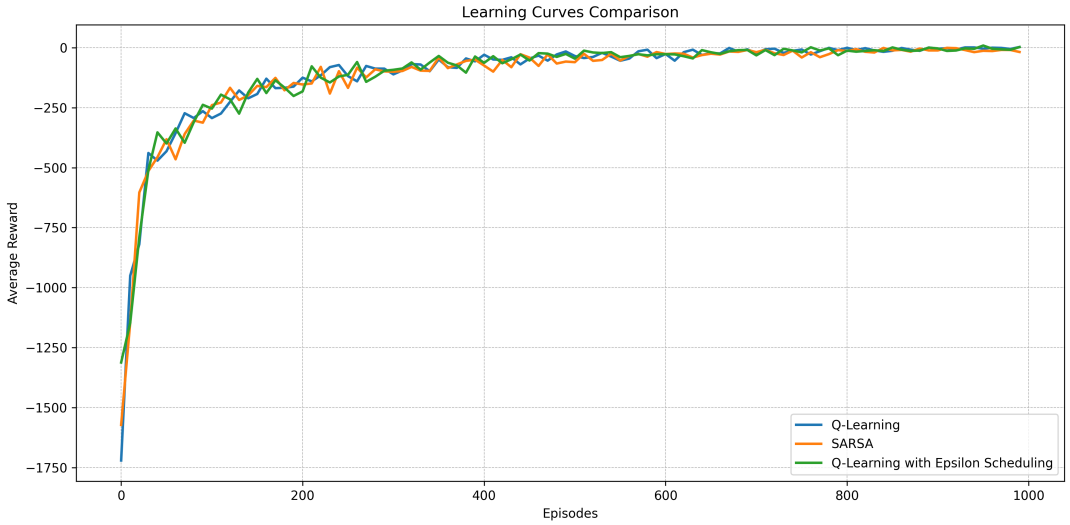

Figure 1: Learning Curves Comparison for Q-Learning, SARSA, and Q-Learning with Epsilon Scheduling