CS7641 A4: Markov Decision Processes

Scott Schmidl
sschmidl3@gatech.edu

I. Introduction

Reinforcement learning (RL) stands as a cornerstone methodology for enabling agents to learn optimal behaviors within a specified environment through trial and error. Among the various algorithms that facilitate this learning process, Q-learning, value iteration, and policy iteration are prominent for their efficacy and versatility. This paper delves into the application of these algorithms within two distinct environments: the grid-world problem known as Frozen Lake and non-grid-world problem known as Blackjack. Both environments present unique challenges, making them nice for evaluating the effectiveness of our chosen methods. Through this study, I aim to not only elucidate the nuances of applying Q-learning, value iteration, and policy iteration in varied contexts but also to contribute to the broader understanding of reinforcement learning's potential in solving complex decision-making problems.

Q-learning is a reinforcement learning algorithm that learns optimal action-selection policies for Markov decision processes. It iteratively updates a Q-table, representing the expected cumulative rewards for taking each action in each state, by exploring the environment and using the Bellman equation to update Q-values towards the optimal policy.

Value iteration is a dynamic programming algorithm used to solve Markov decision processes by iteratively computing and improving the value function until convergence. It involves estimating the value of each state by considering the expected cumulative rewards achievable from that state onward, converging to the optimal value function and policy.

Policy iteration is an iterative algorithm for finding the optimal policy in Markov decision processes. It alternates between two steps: policy evaluation, where the value function for a given policy is computed, and policy improvement, where the policy is updated based on the current value function. This process repeats until convergence to the optimal policy.

II. First MDP

Quisque faucibus egestas fermentum.

A. Description

The 16x16 Frozen Lake problem is an extension of the classic Frozen Lake problem in reinforcement learning. In this grid-world environment, the agent navigates a grid of size 16x16, encountering frozen (with a 90% chance) and hole tiles. The goal is to reach the designated goal tile while avoiding falling into holes. Actions include moving up, down, left, and right. The challenge lies in the stochastic nature of the ice, where the agent might slide in a direction different from the intended one with a probability of $\frac{1}{3}$. I also explore a 4x4 Frozen Lake problem as a comparison.

B. Results

Pellentesque efficitur magna pharetra, molestie libero vel, tempus justo.

1) Policy-Iteration: Pellentesque auctor eros justo, nec cursus ligula porta tincidunt.

2) Value-Iteration: Nulla pharetra felis ut felis auctor convallis.

3) Q-Learning Algorithm: Morbi porttitor mi neque, at sollicitudin odio imperdiet ut.

C. Summary

Praesent in scelerisque mauris.

III. Second MDP

Quisque faucibus egestas fermentum.

A. Description

Blackjack is a popular card game often used as a baseline in reinforcement learning algorithms. In this game, the player competes against the dealer, aiming to obtain a hand value as close to 21 as possible without exceeding it. The player can choose to "hit" or "stand" based on their current hand value and the visible card of the dealer. The dealer follows a fixed strategy, usually standing on a hand value of 17 or higher. The challenge in reinforcement learning lies in learning the optimal policy for maximizing long-term rewards while taking into account the hidden information about the dealer's cards.

B. Results

Pellentesque efficitur magna pharetra, molestie libero vel, tempus justo.

1) Policy-Iteration: Pellentesque auctor eros justo, nec cursus ligula porta tincidunt.

2) Value-Iteration: Nulla pharetra felis ut felis auctor convallis.

3) Q-Learning Algorithm: Morbi porttitor mi neque, at sollicitudin odio imperdiet ut.

C. Summary

Praesent in scelerisque mauris.

IV. Conclusion

This investigation into the application of Q-learning, value iteration, and policy iteration within the Frozen Lake and Blackjack environments displays the adaptability and potential of reinforcement learning algorithms in navigating complex spaces. My findings reveal that while no single algorithm uniformly outperforms the others across all metrics and environments, each possesses distinct characteristics that may render it more suitable to specific types of problems. For instance, Q-learning's model-free approach provides flexibility in unstructured environments, whereas value iteration and policy iteration offer efficiency and stability in environments where the model can be accurately defined. These insights contributed to my understanding of reinforcement learning algorithms but also have practical implications for their application in diverse fields.

V. Resources

[1] Nakamura, K. (2023). *CS7641 ML - A4 Template*.
[2] Mansfield, John. (2024). *BetterMDPTools* https://github.com/jlm429/bettermdptools.