Gambling using Reinforcement Learning

UID - 74

Rahul Sharma (23B2528)

February 14, 2025

Abstract

Reinforcement Learning (RL) is a powerful paradigm in artificial intelligence, widely used for solving complex decision-making problems. This report chronicles a structured learning journey in RL, covering fundamental concepts, practical implementations, and real-world applications. The project spans multiple weeks, each dedicated to mastering specific aspects of RL, culminating in a hands-on Blackjack simulation. This exploration has not only deepened theoretical understanding but also reinforced implementation skills through progressive assignments.

Progress Summary

Week 0: Introduction to Python and Tools

The first week served as an introductory phase where I familiarized myself with Python's syntax and essential libraries such as NumPy and Matplotlib. These libraries are fundamental for numerical computing and visualization, which are critical for understanding RL models. Additionally, I explored Jupyter Notebook, a powerful interactive environment that facilitates coding, visualization, and debugging in an efficient manner.

Task Completed: Week 0 Assignment

Week 1: Fundamentals of Markov Decision Processes (MDPs)

Week 1 was dedicated to understanding Markov Decision Processes (MDPs), which form the backbone of RL. MDPs provide a mathematical framework for modeling decision-making problems where outcomes are partly random and partly controlled by an agent. I studied state transitions, rewards, and policies, implementing models for Bandit Walk, Slippery Bandit Walk, and Frozen Lake environments to gain practical exposure.

Task Completed: Week 1 Assignment

Week 2: Policy Optimization and Value Estimation

Building upon the foundation of MDPs, Week 2 focused on optimizing policies to maximize rewards. I explored concepts such as value functions, policy improvement, and iterative methods for policy evaluation. The key takeaway was the understanding and implementation of two core algorithms: policy iteration and value iteration. These techniques enable finding optimal policies that dictate the best actions for each state. I applied these methodologies to the Frozen Lake MDP scenario, analyzing how different strategies affected outcomes.

Task Completed: Week 2 Assignment

Week 3: Multi-Armed Bandits and Exploration-Exploitation Trade-off

In Week 3, I explored the Multi-Armed Bandit (MAB) problem, a simpler form of RL where an agent repeatedly selects from a set of actions to maximize total reward. The challenge in MAB lies in balancing exploration (gathering more information) with exploitation (choosing the best-known option). I implemented several approaches to solve MABs, including:

- **Epsilon-Greedy Strategy**: A simple method that explores randomly with probability ε and exploits the best-known action otherwise.
- **Upper Confidence Bound (UCB)**: An approach that prioritizes actions with uncertain reward estimates, encouraging strategic exploration.
- **KL-UCB**: A variant of UCB that incorporates Kullback-Leibler divergence to fine-tune exploration.
- **Thompson Sampling**: A Bayesian approach that samples from probability distributions to make better decisions over time.

These implementations deepened my understanding of the exploration-exploitation dilemma and its impact on decision-making.

Task Completed: Week 3 Assignment

Week 4: Real-World Application - Blackjack Simulation

Week 4 marked the application phase, where I implemented a reinforcement learning model for the popular card game Blackjack. The goal was to optimize decision-making strategies using TD(λ), an approach that bridges the gap between Monte Carlo (MC) methods and Temporal Difference (TD) learning. The model estimated state-value functions for various observation tuples, allowing the agent to make more informed decisions.

Key concepts applied:

- **Monte Carlo Methods**: Estimating returns based on complete episodes.
- **Temporal Difference Learning**: Updating value functions incrementally without waiting for an episode to end.
- **TD(λ)**: A hybrid approach that balances the benefits of MC and TD learning by adjusting the λ parameter.

This project solidified my understanding of how RL can be utilized in real-world decision-making scenarios.

Task Completed: Blackjack Project

Conclusion

This structured exploration of reinforcement learning has been instrumental in developing both theoretical knowledge and practical implementation skills. The progressive approach, starting from fundamental concepts and gradually advancing to real-world applications, has provided a comprehensive understanding of RL techniques. By implementing multiple RL strategies and evaluating their effectiveness, I have gained valuable insights into designing intelligent decision-making models.

Future directions include delving deeper into deep reinforcement learning, exploring neural network-based policy approximations, and experimenting with RL in more complex environments. The knowledge acquired through this project lays a strong foundation for further research and applications in AI-driven decision-making problems.