Akshar Zala - Report

WiDS: A Journey Through Reinforcement Learning

Introduction

I attended WiDS 2024 at IIT Bombay's Analytics Club. My project, "A Journey through Reinforcement Learning," was guided by Balaji Karedla and Nirav Bhattad. This project was both thrilling and educational. The realm of AI has always fascinated me, and I have often wondered how AI learns. This project provided the answers to many of my questions. I studied Reinforcement Learning (RL) and was able to progress from a beginner to an advanced level, gaining both theoretical knowledge and hands-on experience. I implemented a 15-Puzzle Solver, applying everything I had learned throughout the earlier weeks of the project. This report captures my learning experience, key takeaways, and how I plan to apply this knowledge in the future.

Foundation Building (Week 0)

Python and Fundamental Libraries

Having already learned C++ in the CS101 course, transitioning to Python was a different and thrilling adventure. I discovered that Python has powerful libraries that simplify problem-solving. The key libraries I focused on were:

- NumPy – Efficient numerical computation
- Pandas – Data organization and analysis
- Matplotlib – Effective data visualization

Mastering these tools was like discovering new approaches to solving problems, providing a solid foundation for the RL concepts that followed.

Version Control with Git

Learning Git was another crucial milestone. Initially, it seemed daunting, but as the project progressed, I truly came to appreciate it. The ability to:

- Track changes in code
- Manage multiple versions efficiently

- Collaborate effectively with teammates

made a tangible impact on my workflow.

Multi-Armed Bandit Problem (Week 1)

The first deep dive into RL began with the Multi-Armed Bandit problem, as presented in Sutton and Barto's book. This fundamental problem laid the groundwork for understanding how agents learn to act optimally in uncertain environments.

Key Implementations

- Implemented the greedy algorithm, a naive but straightforward approach.
- Implemented the ε-greedy algorithm, balancing exploration and exploitation.
- Compared both approaches and analyzed their performance.

Learning Outcomes

This hands-on learning experience was a revelation. I observed how the ε-greedy policy significantly improved decision-making by balancing exploration of new opportunities with exploitation of known rewards. This theme was recurrent throughout RL.

Markov Decision Processes (Week 2)

This week focused on the mathematical backbone of RL: Markov Decision Processes (MDPs). Initially, the theory felt abstract, but modeling simple environments helped me grasp how MDPs codify the state-action-reward framework essential to RL.

Fundamental Concepts Gained

- State transition and reward schemes in MDPs.
- Markov chains and their probabilistic nature.
- Expected values and random variables to simulate decision-making.

It was fascinating to learn that most practical RL applications—ranging from game AI to robotics—utilize some form of MDP, even when not explicitly formulated.

Dynamic Programming and MDP Applications (Week 3)

This week bridged theory and practice, focusing on solving dynamic programming (DP) problems in MDPs.

Key Implementations

- Implemented policy evaluation and policy iteration techniques.
- Solved DP problems using various optimization methods.
- Analyzed how slight policy modifications impact long-term decision-making.

This was the turning point—the connection between MDPs, DP, and RL became evident, linking theoretical concepts with real-world applications.

Final Project: 15-Puzzle Solver

With weeks of learning behind me, I approached the 15-puzzle, a well-known sliding tile puzzle that tests computational efficiency and solution strategies.

Implementation Details

1. State Space Representation

- Designed an efficient board representation to work with puzzle layouts.
- Implemented state transition functions to define valid movements.
- Developed a solvability checker to identify impossible puzzles.

2. Search Algorithm

- Applied RL techniques to enhance puzzle-solving strategies.
- Experimented with heuristic search and optimization methods.
- Ensured efficient pathfinding to reduce computational costs.

3. Key Features

- Automatic unsolvable configuration detection, preventing unnecessary computation.
- Optimized solution path construction using RL insights.
- Performance improvements to enhance runtime efficiency.

This project was more than just an implementation—it was a culmination of everything I had learned. Witnessing the solver successfully solve puzzles was incredibly rewarding and reaffirmed my confidence in translating RL concepts into real-world problems.

Applications and Future Directions

Beyond the 15-puzzle, the concepts explored in this project have vast applications in AI and decision-making systems. Some key areas where RL is transformative include:

- Game AI – From chess engines to adaptive in-game opponents.
- Robotics – Path planning and autonomous navigation.
- Optimization Problems – Supply chain logistics, resource allocation, and more.
- Finance and Trading – Reinforcement learning-driven stock market strategies.

This project has ignited my passion to further explore these areas, extending RL applications beyond academic exercises into real-world, impactful solutions.

Personal Growth

Reflecting on this experience, the WiDS-2024 project has been a transformative learning journey. Key takeaways include:

- Mastery of RL concepts and their practical applications.
- Enhanced problem-solving skills, breaking down complex problems into manageable steps.
- Stronger programming abilities, improving code structuring and efficiency.
- Exposure to technical documentation, and learning to explain complex ideas clearly.

Progressing from Python fundamentals to developing an advanced puzzle solver was challenging yet rewarding. Every obstacle was an opportunity to think critically, refine my approach, and grow as a learner.

Acknowledgments

I am deeply grateful to Nirav Bhattad and Balaji Karedla for their invaluable guidance throughout this project. Their structured mentorship and enthusiasm made Reinforcement Learning a fun and fulfilling experience. This journey would not have been as enriching without their thought-provoking discussions and encouragement to experiment and explore.