EvolveRL - Reinforcing Natural Selection

A THIRD YEAR PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF B.Sc. IN COMPUTATIONAL MATHEMATICS

BY

1. SUPRIYA BANIYA (028295-20)

2. AYAM BASYAL (028298-20)

3. SHAMBHAV RAYAMAJHI(028311-20)

4. YOGESH SAPKOTA(028312-20)

.5 RUBY SHRESTHA(028316-20)

SCHOOL OF SCIENCE

KATHMANDU UNIVERSITY

DHULIKHEL, NEPAL

JANUARY 2024

DECLARATION

We “Ayam Basyal”, ”Supriya Baniya”, ”Ruby Shrestha”, ”Yogesh Sapkota” and ”Shambhav Rayamajhi” hereby declare that the work contained herein is entirely our own, except where states otherwise by reference or acknowledgment, and has not been published or submitted elsewhere, in whole or in part, for the requirement for any other degree or professional qualification. Any literature, data or works done by others and cited within this dissertation has been given due acknowledgment and listed in the reference section.

Ayam Basyal	**Supriya Baniya**	**Ruby Shrestha**
Registration no. : 028298-20	Registration no. : 028295-20	Registration no. : 028316-20
Date:	Date:	Date:

Yogesh Sapkota	**Shambhav Rayamajhi**
Registration no. : 028312-20	Registration no. : 028311-20
Date:	Date:

CERTIFICATION

This project entitled "EvolveRL - Reinforcing Natural Selection" is carried out under my supervision for the specified entire period satisfactorily, and is hereby certified as a work done by following students:

1. SUPRIYA BANIYA (028295-20)
2. AYAM BASYAL (028298-20)
3. SHAMBHAV RAYAMAJHI(028311-20)
4. YOGESH SAPKOTA(028312-20)

.5 RUBY SHRESTHA(028316-20)

in partial fulfillment of the requirements for the degree of B. Sc. in Computational Mathematics, Department of Mathematics, Kathmandu University, Dhulikhel, Nepal.

Mr. Harish Chandra Bhandari
Department of Mathematics,
School of Science, Kathmandu University,
Dhulikhel, Kavre, Nepal
Date:

APPROVED BY:

I hereby declare that the candidate qualifies to submit this report of the Semester Project (COMP 311) to the Department of Mathematics.

Prof. Dr. Rabindra Kayastha
Head of the Department
Department of Mathematics
School of Science
Kathmandu University
Date:

ACKNOWLEDGMENTS

First and foremost, we extend our deepest appreciation to our project supervisor, Mr. Harish Chandra Bhandari for his guidance, expertise, and unwavering support throughout the entire duration of this project. His valuable insights and constructive feedback has greatly shaped the direction and quality of our work. We are also grateful to the faculty members of the Department of Mathematics at Kathmandu University for their valuable input, encouragement, and assistance. Their expertise and academic guidance have been invaluable in expanding our understanding and enhancing the overall project. We would like to acknowledge the contributions of our fellow team members and colleagues who have actively participated in the project. Their collaboration, dedication, and shared knowledge have been instrumental in accomplishing the project milestones and overcoming challenges. We would like to collectively acknowledge and thank all those mentioned above, as well as any other individuals who may have contributed in ways both seen and unseen. Your support, guidance, and contributions have been crucial in the successful completion of this project.

ABSTRACT

EvolveRL is an innovative project that combines reinforcement learning and predator-prey dynamics to get to the heart of evolution. In this simulated environment, digital agents assume the roles of predator and prey and navigate a dynamic world governed by speed and energy parameters. The main goal of this project is to decipher the algorithms that shape adaptive behavior and explore their broad applications. Navigating EvolveRL's simulated landscapes reveals fascinating insights into the subtle complexities of predator-prey dynamics. Digital agents demonstrate adaptive strategies and novel behaviors provide insight into the dynamic interactions within the simulated ecosystem. This project not only contributes to evolutionary biology, but also investigates the potential application of these findings in real-world scenarios. EvolveRL goes beyond scientific exploration, bridging biology and artificial intelligence to provide a digital lab for understanding the subtle workings of evolution. In this digital adventure, researchers, educators, and enthusiasts are invited to join us in uncovering the algorithms that underlie life's continuous adaptation and survival.

keywords : *Optimal utilization; reinforcement learning; Predator and prey agents; Adaptive behaviors.*

CONTENTS

DECLARATION ii

CERTIFICATION iii

ACKNOWLEDGEMENTS iv

ABSTRACT v

LIST OF FIGURES viii

LIST OF SYMBOLS ix

1 INTRODUCTION 1

1.1 **Reinforcement learning** . . . 1

1.2 **Multi-Agent Reinforcement learning** . . . 1

1.3 **Natural Selection** . . . 2

1.4 **Introduction to the Topic** . . . 2

1.4.1 **Literature Review** . . . 2

1.4.2 **Problem Statement** . . . 3

1.4.3 **Motivation** . . . 4

1.5 **Objectives** . . . 5

1.6 **Significance** . . . 5

1.7 **Limitations** . . . 6

2 METHODOLOGY/MODEL EQUATION 7

2.1 **Theoretical/Conceptual Framework** . . . 7

2.2 **The Environment** . . . 9

2.3 **Lotka-Volterra models of Predator-Prey Relationships** . . . 12

3 RESULTS AND DISCUSSIONS 14

4 CONCLUSIONS **16**

LIST OF FIGURES

2.1 Markov Decision Process 7
2.2 The AEC diagram 10
2.3 A hexagonal grid environment in which predators and prey can interact with each other 11

3.1 Lotka Volterra Model after Reinforcement Learning 15

LIST OF SYMBOLS

In a Markov Decision Process:

s, s'	states
a	an action
r	a reward
$q_*(s, a)$	Value of action a taken from state s
R_s^a	Reward gained from taking action a from state s
$P_{ss'}^a$	Probability of transitioning to state s' from s when taking an action a
$v_*(s')$	the value of the next state s' - which is initially assumed

CHAPTER 1

INTRODUCTION

1.1 Reinforcement learning

"Reinforcement learning is a computational approach to understanding and automating goal-directed learning and decision making. It is distinguished from other computational approaches by its emphasis on learning by an agent from direct interaction with its environment, without requiring exemplary supervision or complete models of the environment" [7]. The agent learns through trial and error by receiving feedback in the form of rewards or punishments. The goal of reinforcement learning is for the agent to learn an optimal policy or strategy that maximizes the cumulative reward over time. It involves making sequential decisions, taking into account the current state of the environment, the available actions, and the expected future rewards.

1.2 Multi-Agent Reinforcement learning

"Multi-agent Reinforcement Learning is a computational approach in which multiple agents learn and make decisions in a coordinated manner through reinforcement learning techniques" [4]. In multi-agent reinforcement learning each agent has its own policy and learns from its own experiences as well as those of the other agents. The agents must learn to take into account the actions and policies of the other agents in order to make effective decisions.

1.3 Natural Selection

"Natural selection is the process through which populations of living organisms adapt and change. Individuals in a population are naturally variable, meaning that they are all different in some ways. This variation means that some individuals have traits better suited to the environment than others. Individuals with adaptive traits—traits that give them some advantage—are more likely to survive and reproduce. These individuals then pass the adaptive traits on to their offspring. Over time, these advantageous traits become more common in the population. Through this process of natural selection, favorable traits are transmitted through generations.

Natural selection can lead to speciation, where one species gives rise to a new and distinctly different species. It is one of the processes that drives evolution and helps to explain the diversity of life on Earth."[5]

1.4 Introduction to the Topic

EvolveRL" aims to simulate a dynamic environment where agents, acting as predators and prey, engage in a virtual struggle for survival. The project leverages reinforcement learning, a type of machine learning, to enable these agents to adapt and evolve over time based on their interactions with the environment. The simulated world reflects the complexities of nature, incorporating key factors like speed and energy that influence the agents' behaviors.

This exploration into predator vs. prey dynamics provides valuable insights into adaptive behaviors and the underlying principles of evolution. The project serves as a bridge between the intricacies of nature and the capabilities of artificial intelligence in the digital age. By studying how agents navigate and survive in this simulated environment, the researchers behind "EvolveRL" aim to deepen our understanding of the fundamental principles that govern life and evolution.

1.4.1 Literature Review

Khadka et al.(2020) Evolutionary Reinforcement Learning for Sample-Efficient Multiagent Coordination MERL is a training platform for multi-agent systems that splits the optimization of specific agent rewards and team-based rewards into two separate processes. It uses an evolutionary algorithm to optimize team objectives through neuroevolution

and employs a gradient-based optimizer to maximize individual agent rewards. To share knowledge between the two processes, gradient-based policies are periodically incorporated into the evolutionary population. This approach allows the evolutionary algorithm to leverage skills acquired through agent-specific rewards to improve overall team performance.

Hallawa et al. (2020). Evo-RL: Evolutionary-Driven Reinforcement Learning. In this work, they propose a hybrid approach combining Evolutionary Algorithms(EAs) and reinforcement learning(RL). Most importantly, their approach handles problems where the reward function is not available in many of the state in the state space.

However, for evaluation purposes, we have implemented the algorithm as follows. Firstly, we used the EA in the form of Genetic Programming (GP). However, as for what concerns the representation of the instinctive (evolved) behaviour, we adopted behaviour trees (BTs). These fit well with GP and, unlike ANN, are much easier to interpret. As for the learned behaviour, we adopted two possibilities, one tabular representation used when testing our approach with Q-learning, and another ANN representation when testing our approach with Proximal Policy Optimization (PPO) and Deep Q-Network (DQN) algorithm.

The presented hybrid approach REAL outperforms EA only and RL only, even when adopting state-of-the-art RL algorithms such as PPO and DQN. Our approach has a number of points of strength: it integrates RL in an EA framework, hence benefiting from both methodologies. Furthermore, it works with any RL algorithm and can handle problems where the reward function is not valid in all states.

Jeongho Park et al.(2021), in his research uses multi-agent reinforcement learning to simulate and study the co-evolution of predators and prey in ecosystems. This research aims to develop ecologically plausible behaviors for both predators and prey. The results show that the learned policies lead to sustainable outcomes, reducing the risk of extinction. This approach connects ecological theory with reinforcement learning, providing insights into predator-prey interactions.

1.4.2 Problem Statement

It is intrinsically difficult to comprehend evolution, adaptive behaviors, and the intricate dynamics of predator-prey relationships in the natural world. Conventional approaches find it difficult to describe the complex relationships and ever-changing mechanisms that

shape survival tactics. Because of this, there is a gap in our knowledge of the underlying algorithms that drive evolution. The goal of this study is to fill the gap in the literature by providing a thorough and in-depth analysis of adaptive behavior in predator-prey dynamics. Developing a framework that utilizes the concepts of reinforcement learning to model and examine these intricate relationships is the difficult part. How can the complex dance of survival in a dynamic environment, where agents change and adapt over time, be properly modelled?

By creating and deploying "EvolveRL," a simulated environment that incorporates reinforcement learning concepts into the investigation of predator-prey dynamics, we hope to close this gap. By this project, we hope to advance knowledge of the algorithms controlling adaptation and survival as well as the wider ramifications for the nexus between artificial intelligence and the natural world.

1.4.3 Motivation

The motivation behind our project is to address how different agents, like predators and prey, learn and adapt in a simulated world using Reinforcement learning. Our curiosity was deeply rooted in our fascination with the fundamental principles of nature, particularly the instinct for survival that drives all living beings. Our journey led us to the captivating realm of predator vs. prey dynamics, where agents learn the art of survival through reinforcement learning. These agents, representing predators and prey, navigate a world governed by parameters like speed and energy all while engaging in a primal battle for survival.

We sought to uncover the intricacies of adaptive behavior, evolution, and the delicate balance of nature, all through the lens of reinforcement learning. By simulating these primal struggles, we hoped to unravel the algorithms that underpin survival and adaptation. Through the exploration of predator-prey dynamics and reinforcement learning, we embarked on a path that promised to expand our understanding of the natural phenomenon that is evolution. We combined these ideas to create 'EvolveRL'. It's like a computer game where we simulate how systems change. We add different characters, like predators and prey, Overall, we want to use these combined ideas to learn how the agents evolve and adapt.

1.5 Objectives

Followings are the objectives of the project:

1. Seamlessly integrate machine learning, mathematical principles, and biological phenomena to model adaptive systems realistically.

2. Develop optimal reward and penalty functions for reinforcement learning to drive the evolution of intelligent agents.

3. Simulate predator and prey interactions, mirroring nature's dynamics, to capture nuances of adaptive behaviors.

4. Compare outcomes of real-world natural selection with our model, analyzing emergent behaviors and population dynamics.

1.6 Significance

EvolveRL - Reinforcing Natural Selection carries several significant implications.

1. Advancing Evolutionary Biology: By simulating predator-prey dynamics with reinforcement learning, the project offers a unique lens through which to observe and understand the adaptive behaviors that drive evolution. This insight contributes to the field of evolutionary biology, providing a digital laboratory for studying complex ecological interactions.

2. Unveiling Algorithms of Adaptive Behavior: The project's goal is to identify the fundamental algorithms that control adaptive behaviors in changing surroundings. This advances our knowledge of natural selection and has ramifications for creating artificial intelligence models that are more complex and capable of adapting to changing environments.

3. Insights into Ecosystem Resilience: Gaining knowledge on how predators and prey adjust in artificial habitats can help us understand how resilient ecosystems are. This information can be extremely helpful in forecasting and reducing the effects of environmental changes, which can guide conservation initiatives and ecosystem management plans.

4. Educational Value: The project, presented as the simulated environment "EvolveRL," can serve as an educational tool. It provides a visually engaging and interactive way for students and researchers to explore fundamental concepts in evolution, ecology, and reinforcement learning.

5. Interdisciplinary Bridge: By integrating principles of biology and artificial intelligence, this project demonstrates the potential for interdisciplinary collaboration. It bridges the natural sciences and computational approaches, promoting a more comprehensive understanding of complex ecological phenomena.

In summary, the significance of this project is that it will deepen our understanding of evolutionary processes, contribute to interdisciplinary knowledge, and impact both scientific research and practical applications in fields ranging from conservation to the development of artificial intelligence. It's possible.

1.7 Limitations

1. Linear Functional Response: The model assumes a linear functional response, implying a constant rate of predation per predator. In real ecosystems, the functional response may be more complex, with saturation effects and handling time influencing predation rates.

2. No Immigration or Emigration: The Lotka-Volterra model does not consider immigration or emigration of individuals between populations. In natural ecosystems, the movement of individuals between different areas can have a significant impact on population dynamics.

3. Constant Prey Density: The model assumes that prey are evenly distributed throughout the environment, providing a constant prey density. In reality, prey distribution is often patchy, and spatial heterogeneity can influence predator-prey interactions.

4. Deterministic Nature: The model is deterministic and does not account for stochastic (random) events that can influence population dynamics. Natural systems are subject to uncertainties, such as environmental fluctuations and demographic stochasticity, which are not considered in the basic model.

CHAPTER 2

METHODOLOGY/MODEL EQUATION

2.1 Theoretical/Conceptual Framework

There are certain fundamental mathematical concepts behind Reinforcement Learning. The most basic and the backbone concept is the Markov Decision Process [7]. A Markov Decision Process (MDP) has an environment, an agent on the environment, different states(S), actions(A), and the reward(R).

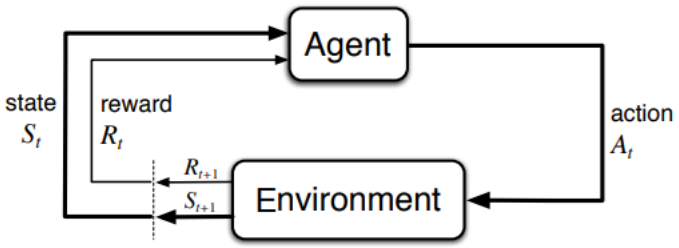

Figure 2.1: Markov Decision Process

In a finite Markov Decision Process (MDP), the states (S), actions actions(A), and rewards (R) are all limited in number, allowing the random variables R_t and S_t to have clear discrete probability distributions that rely solely on the previous state and action [7]. We create state value function to determine which state is better than the other for the agent. Similarly, we also generate action-value function to determine rewards for a certain action taken by the agent. Some terminologies related to our project's initial problem are defined below:

- Policy: In reinforcement learning, a policy is a strategy or a mapping from states to actions that guides the behavior of an agent in an environment. It defines the agent's decision-making process, specifying which action to take in each state [1].

- Action-Value Function: In reinforcement learning, the action-value function, also known as the Q-function, is a critical component used to estimate the value of taking a particular action in a given state. Following mathematical expression [7] represents the action-value function:

$$q_*(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_*(s') \tag{1}$$

where,
$q_*(s,a)$ = Value of action a taken from state s ,
R_s^a = Reward gained from taking action a from state s,
$P_{ss'}^a$ = Probability of transitioning to state s' from s when taking an action a,
$v_*(s')$ = the value of the next state s' - which is initially assumed

- State-Value Function: A state value function, denoted as $v_*(s)$, is a fundamental concept in reinforcement learning and is used to estimate the expected cumulative reward an agent can achieve from being in a particular state s and following a given policy π [3].

$$v_*(s') = max\ q_*(s,a) \tag{2}$$

In order to solve a reinforcement learning problem like ours, Bellman Equation is very crucial. The Bellman Equation below helps to calculate State-Action pair under a policy π.

$$\underbrace{\text{New}q(s,a)}_{\text{New Q-Value}} = q(s,a) + \underbrace{\alpha}_{\text{Learning rate}}[\underbrace{R(s,a)}_{\text{Reward}} + \underbrace{\gamma}_{\text{Discount rate}} \overbrace{\max q^*(s',a')}^{\substack{\text{Maximum predicted reward, given} \\ \text{new state and all possible actions}}} - q(s,a) \tag{3}$$

Another important concept surrounding our project is a discount factor. A discount factor is a constant parameter that relates reward to time or number of iterations. It is a

number between 0 and 1. It helps to scale down future rewards and prioritize the initial reward. Empirical studies have shown the possibility of significantly reducing the number of learning steps using an appropriate discount factor. [2]

A fundamental concept in reinforcement learning which has seen existence since the inception of the reinforcement learning concept is a Monte Carlo simulation. In reinforcement learning, Monte Carlo simulation is a technique used to estimate the value of states or state-action pairs based on sampling and averaging returns from simulated episodes. It involves repeatedly running the agent's policy in an environment, collecting experience, and updating the value function based on the observed returns.
By sampling multiple trajectories and averaging the returns, Monte Carlo simulation can provide unbiased estimates of the state or state-action values, allowing the agent to learn optimal policies and make informed decisions in a stochastic and uncertain environment. John von Neumann and Stanislaw Ulam developed the Monte Carlo Method during World War II with the goal of enhancing decision-making in situations characterized by uncertainty. [6] The method acquired its name from the famous gambling destination, Monaco, due to its reliance on chance-based modeling techniques, akin to playing roulette.

2.2 The Environment

In an AEC(Agent environment cycle) environment, agents act sequentially, receiving updated observations and rewards before taking an action. The environment updates after each agent's step, making it a natural way of representing sequential games such as Chess. The AEC model is flexible enough to handle any type of game that multi-agent RL can consider with the underlying environment updating after each agent's step. Agents receive updated observations and rewards at the beginning of their actions.

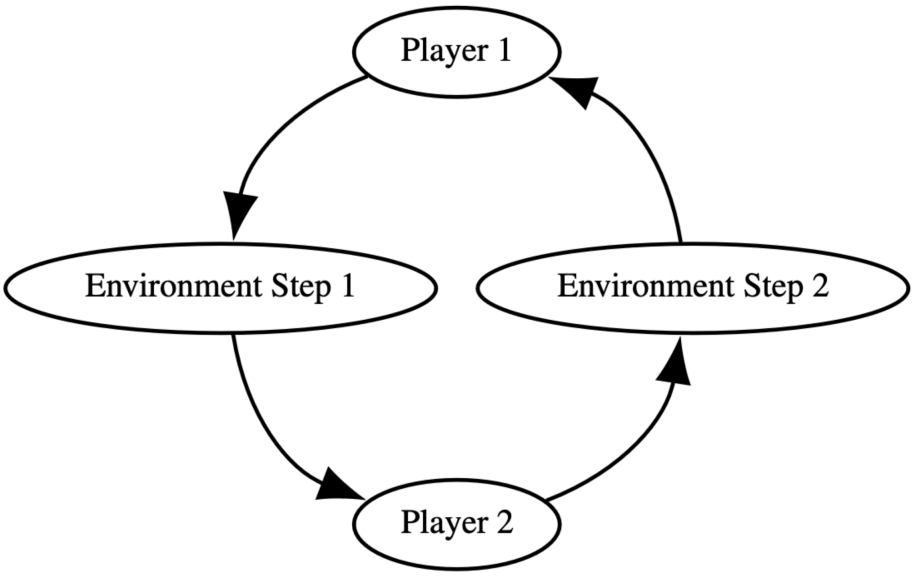

Figure 2.2: The AEC diagram

This is in contrast to the Partially Observable Stochastic Game (POSG) model, represented in our Parallel API, where agents act simultaneously and can only receive observations and rewards at the end of a cycle. This makes it difficult to represent sequential games, and results in race conditions–where agents choose to take actions which are mutually exclusive. This causes environment behavior to differ depending on internal resolution of agent order, resulting in hard-to-detect bugs if even a single race condition is not caught and handled by the environment (e.g., through tie-breaking).

The AEC model is similar to Extensive Form Games (EFGs) model, used in DeepMind's OpenSpiel. EFGs represent sequential games as trees, explicitly representing every possible sequence of actions as a root to leaf path in the tree. A limitation of EFGs is that the formal definition is specific to game-theory, and only allows rewards at the end of a game, whereas in RL, learning often requires frequent rewards.

Our Environment

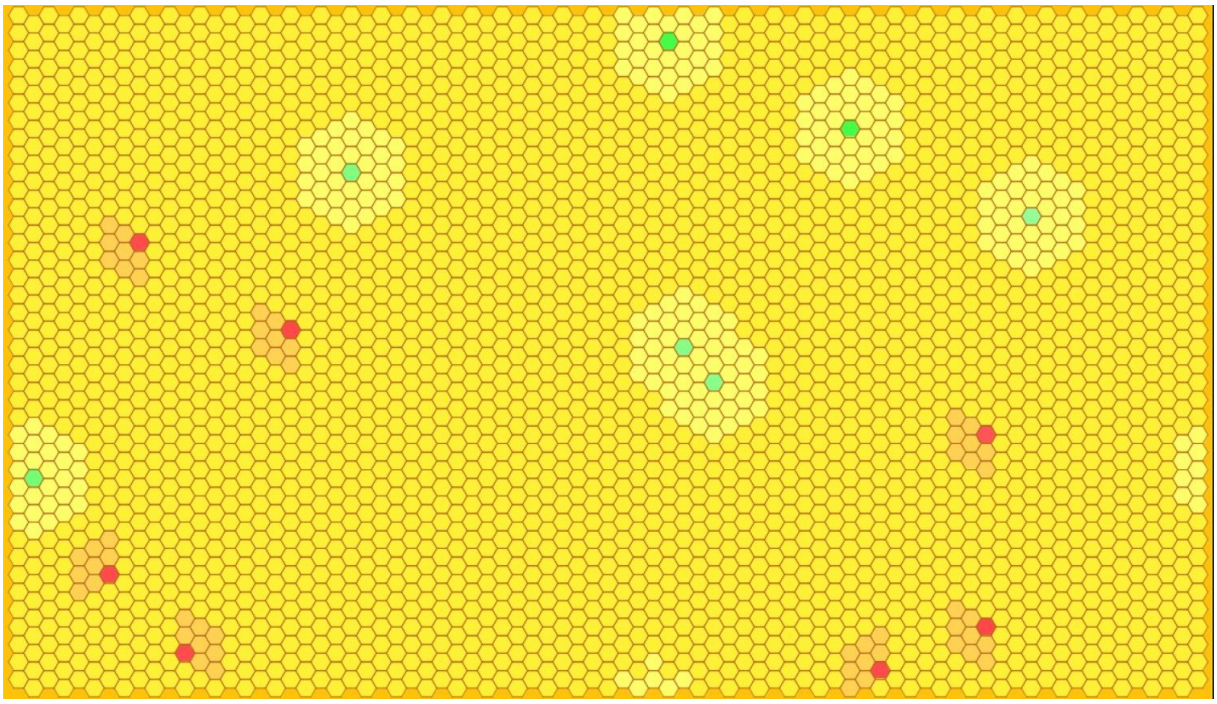

Figure 2.3: A hexagonal grid environment in which predators and prey can interact with each other

In our hexagonal grid environment, comprised of 3081 hexagons configured in a 79x39 layout, efficient traversal of agents within the grid poses a significant challenge. Recognizing the complexity inherent in navigating a hexagonal grid, we undertook the task of optimizing our data representation. This led us to adopt the axial coordinate system,To address the challenges associated with traversing agents within our hexagonal grid, we seamlessly converted the hexagon storage from a list-based structure to the axial coordinate system. This transformation not only streamlined our data representation but also facilitated a more organized and intuitive approach to grid manipulation. Each hexagon in the list is now indexed by its axial coordinates, rendering navigation more transparent and enabling efficient agent movement within the grid. a strategic move aimed at simplifying grid navigation and enhancing overall efficiency.To address the challenges associated with traversing agents within our hexagonal grid, we seamlessly converted the hexagon storage from a list-based structure to the axial coordinate system. This transformation not only streamlined our data representation but also facilitated a more organized and intuitive approach to grid manipulation. Each hexagon in the list is now indexed by its axial coordinates, rendering navigation more transparent and enabling efficient agent movement within the grid.

2.3 Lotka-Volterra models of Predator-Prey Relationships

The Lotka-Volterra equations were developed to describe the dynamics of biological systems. This system of non-linear differential equations can be described as a more general version of a Kolmogorov model because it focuses only on the predator-prey interactions and ignores competition, disease, and mutualism which the Kolmogorov model includes The Lotka-Volterra equations can be written simply as a system of first-order non-linear ordinary differential equations (ODEs). Since the equations are differential in nature, the solutions are deterministic (no randomness is involved, and the same initial conditions will produce the same outcome), and the time is continuous (the generations of predators and prey are continually overlapping). As with many other mathematical models, many assumptions were made in the creation of the Lotka-Volterra equations. Such assumptions include:

1. There is no shortage of food for the prey population.
2. The amount of food supplied to the prey is directly related to the size of the prey population.
3. The rate of change of population is directly proportional to its size.
4. The environment is constant and genetic adaptation is not assumed to be negligible.
5. Predators will never stop eating.

After such assumptions are made, the Lotka-Volterra equations can be written as:

$$\frac{dx}{dt} = \alpha x - \beta xy$$
$$\frac{dy}{dt} = \delta xy - \gamma y$$

where:

$x =$ number of prey

$y =$ number of predators

$\frac{dx}{dt}$ and $\frac{dy}{dt} =$ the instantaneous rates of the prey and predators, respectively.

$t =$ time

$\alpha, \beta, \delta, \gamma =$ positive real constants

Taking a closer look at the prey equation we can see that the prey are assumed to reproduce exponentially which is represented by the term αx . The equation also shows that the rate at which predators kill prey is proportional to the product of the number of prey and the number of predators, or in other terms, how often the two populations meet. This is represented by the term βxy. Therefore, if there is no population of prey or no population of predators, no decrease in the population of prey (also known as predation) can occur. The equation for prey can be summed up as: the rate at which new prey is born, minus the rate at which prey is killed off. Looking now at the predator equation we can see that the growth of the predator population is proportional to the amount of times the two populations meet. This is similar to the rate at which predators kill prey except that a different constant is used to describe this relationship since the rate at which predators kill and the rate at which they reproduce are not identical. This term is represented by the term δxy. Since prey cannot kill the predators, the decrease in predator population is due to death by natural causes or by emigration. This is assumed to be an exponential decay which is represented by the term γy. The equation for predators can be summed up as: the rate at which they consume prey, minus the natural death rate of the population.[8]

CHAPTER 3

RESULTS AND DISCUSSIONS

In our RL simulation, we observed dynamic interactions between predators and prey over time, capturing cyclical patterns indicative of predator-prey dynamics. The RL model demonstrated adaptability with population oscillations, contrasting with the more deterministic trajectory of the Lotka-Volterra model. Quantitative measures underscored nuanced differences, emphasizing the RL model's sensitivity to changes in initial conditions and environmental factors. This adaptability suggests the potential of RL to capture realistic and dynamic ecological responses.

Comparing RL results with Lotka-Volterra predictions revealed insights into the models' behavior. RL's adaptability surpassed the deterministic nature of Lotka-Volterra, with quicker recovery from disturbances. Although equilibrium points aligned, quantitative differences prompted exploration into underlying mechanisms. These findings suggest RL's flexibility and resilience in simulating predator-prey interactions.

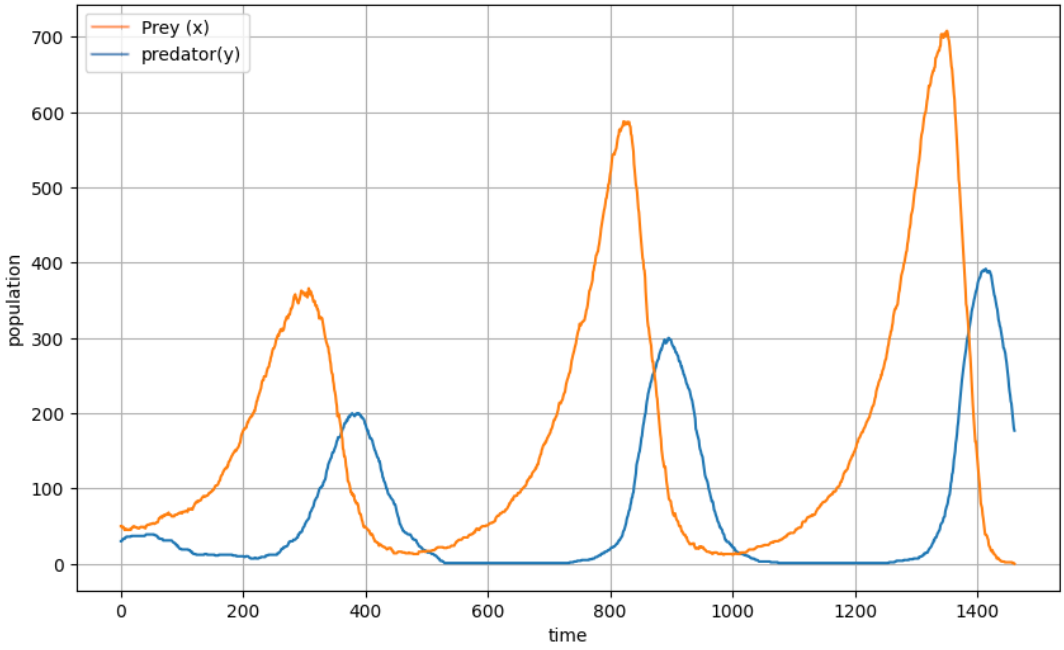

Figure 3.1: Lotka Volterra Model after Reinforcement Learning

In conclusion, our project successfully implemented a predator vs. prey simulation using RL, showcasing the adaptability of the parallel API. Visualizations and comparisons highlighted RL's strengths in adaptive modeling, complementing the deterministic approach of Lotka-Volterra. This study contributes insights into RL's applications in ecological modeling, emphasizing the importance of both data-driven and theoretical approaches for a comprehensive understanding of complex ecological systems.

CHAPTER 4

CONCLUSIONS

In concluding our project, we've reached a major achievement with a successful predator vs. prey simulation, thanks to reinforcement learning (RL). This not only highlights the capabilities of artificial intelligence but also showcases how our intelligent agents adapt, revealing the essence of natural selection.

Our RL framework's well-designed reward and penalty functions enable agents to evolve in response to challenges, mimicking the adaptive dance seen in nature. This aligns smoothly with our first goal of integrating machine learning, math, and biology for a realistic model of adaptive systems. Moreover, our project closely approximates the Lotka-Volterra model, confirming the accuracy of our simulation and validating our approach.

In essence, our project demonstrates the potential blend of AI and ecological principles, offering insights into the intricate dynamics of predator-prey interactions. The fusion of machine learning, math precision, and imitation of natural selection reveals a space where technology mirrors the beauty of the natural world.

In a final reflection, our project underscores the strength of interdisciplinary collaboration, showing how marrying advanced tech with biological insights can lead to innovative solutions. This journey expands the horizons of AI and deepens our appreciation for the wonders of nature, guiding our pursuit of knowledge.

REFERENCES

[1] Dimitri Bertsekas, *Multiagent value iteration algorithms in dynamic programming and reinforcement learning*, Results in Control and Optimization **1** (2020), 100003.

[2] Vincent Francois, Raphael Fonteneau, and Damien Ernst, *How to discount deep reinforcement learning: Towards new dynamic strategies*, (2015).

[3] Yagna Patel, *Optimizing market making using multi-agent reinforcement learning*, (2018).

[4] Hugo Santana, Geber Ramalho, Vincent Corruble, and B. Ratitch, *Multi-agent patrolling with reinforcement learning*, 02 2004, pp. 1122– 1129.

[5] National Geographic Society, *On the origin of species*, pp. 116–251, 05 2015.

[6] Jerome Spanier, Yousry Azmy, and Enrico Sartori, *Monte carlo methods*, pp. 117–165, 04 2010.

[7] Richard S. Sutton and Andrew G. Barto, *Reinforcement learning*, 2nd ed., The MIT Press, 2018.

[8] V. (1928). Volterra, *Variations and fluctuations of the number of individuals in animal species living together*, pp. 3–51, 0 2010.