Reinforcement Learning in Pokémon Red to explore complex multi-reward environments

by

LIAM O'DRISCOLL
URN: 6640106

October 24, 2023

Department of Computer Science
University of Surrey
Guildford, Surrey
England, United Kingdom
GU2 7XH

Project Supervisor: Sotiris Moschoyiannis

I declare that this dissertation is my own work and that the work of others is acknowledged and indicated by explicit references.

Liam O'Driscoll
October 24, 2023

1 Introduction

This section needs to contain an introduction to the problem, aims and objectives (0.5 pages)

[FROM THE HANDBOOK]
It is designed to help you initially choose a project. During the academic year, it also helps you keep track of deliverables; a complete description of each deliverable is provided. Finally, this document specifies how the project is assessed, including pointers to marking criteria, details on the examination, as well as the process for dealing with complaints

1.1 Aims

✖ The aim of this project is to develop a RL agent to play Pokémon Red.

1.2 Objectives

♦ Explore and understand the use of RL techniques in a custom game environment.

♦ Design and implement Pokémon Red game to be a suitable RL environment.

♦ Review and critique literature that applies RL to Pokémon.

♦ Evaluate the performance of different RL algorithms used to train agents within the environment.

♦ Evaluate performance of agents to different forms of reward incentive.

♦ Recommend further developments to the project and applications to real world projects.

1.3 Extension Objectives

♣ Implement different methods to improve the agent's learning and understanding of the environment.

♣ Evaluate performance of the agent when given data in different forms.

♣ Evaluate performance of internal agent to performance of literature.

2 Literature Review

2.1 What other similar work has been done?

2.2 What will I introduce that build son other's work?

3 Technical overview

- Hyperparameter tuning to find optimal performance per experiment.
- Comparison of Gradient Descent and Value based models.
  - Value based:
    - Proximal Policy Optimization
  - Gradient Descent:
    - Actor-Critic Methods: A2C
    - Deep Deterministic Policy Gradient
- Evaluating change in Q values to learning the optimal model using DQN
- Explore the benefit of applying meta learning.

4 Workplan

Month	Goals
October	• Rough structure of the report has been made. • Papers surrounding the project have been read (e.g., similar projects, algorithms that will be explored and technologies to be implemented) • Coding for the project is at its early stages. • Project Synopsis completed and submitted.
Novemebr	• Research and test which algorithms are applicable for comparison and applicable to project. • Draft introduction completed with a basic explanation of RL and how it is suitable for my environment. • Implementation of the Environment is complete
December	• Minimum viable product of code is achieved • Alter reward functions to give different incentives • Problem Analysis has been written • Design documentation and choice has been started
January	• Hyperparameter train sets of agents per algorithm • Train agents on different algorithms • Complete Design choice • Start evaluation of agents
February	• Any necessary extra agent training to be compelted • First version of Report is at a Submittable state
March	• Debugging time for any potential issues • Review of draft report submission
April	• Consider completing Extension Objectives • Final report completed • Time allocated for debugging or potential issues
May	• Last final checks on final version of report • Time allocated for debugging or potential issues