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

by

LIAM O'DRISCOLL
URN: 6640106

October 27, 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 27, 2023

0.1 Definitions

- ♠ Agent: The decision making mechanism recieving state information and performing chosen actions.
- ♠ Environemnt: The world in which the agent interacts with.
- ♠ State: A representation of the environment at the current timestep.
- ♠ Timestep: A value that increments after each action has passed since the start of the episode.
- ♠ Episode: An instance of the environment that the agent is interacting with.
- ♠ Action: The choice made by the agent in response to the state.
- ♠ Reward: The return value when an action is applied to a state.
- ♠ Reward Function: The mechanism in the environment that indicates how well the selected action is to achieving the goal of the environment.
- ♠ Policy: The decision making mechanism within the agent that decides the best action to perform given the state.

1 Introduction

The aim of this project is to develop a reinforcement learning agent to play Pokémon Red to compare the effectiveness of different styles of RL algorithms and their effectiveness. RL is an area of machine learning (ML) where agents make decisions and perform actions on states to achieve a goal.

1.1 Aims

✖ The aim of this project is to develop a RL agent to play Pokémon Red to compare the effectiveness of different styles of RL algorithms and their effectiveness to learn complex reward functions.

1.2 Objectives

- Research applications of RL to Pokemon and conduct a literature review on them
- Implement Pokémon Red game to be a suitable for training of different RL algorithms.
- Evaluate the performance of different RL algorithms used to train agents within the environment.
- Evaluate performance of agents to different forms of rewards functions.
- Recommend further developments to the project and applications to real world projects.

2 Literature Review

In RL, the agent learns through experiences and 'trial and error'. Initially, it has a lack in understanding of the environment. However, through random action selection and the reward that it recieves, it is able to learn an understanding of the environemnt. The agent is incentivized to maximise its reward and will aim to find actions that will yield more reward. This constant state, action and reward loop is what helps the agent improve by altering the policy after every cycle.

RL is different to common traditional machine learning techiniques as it learns "how to map situations to actions-so as to maximize a numerical reward signal." Agent's requirement to learn through experience and actions performed on current states not only affect the present, but also affect future states and actions are two characteristics which distinguishes itself from other forms of ML. It is also what makes Pokémon Red a suitable environment to apply this style of ML to.

[MUST mention how important markov chains are and how they are applicable to pokemon!]

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