Reinforcement Learning in Pursuit-Evasion Games

Abstract

This project presents an approach to the classical pursuit-evasion game problem using reinforcement learning. We tackle the scenario of a single pursuer agent attempting to capture an evading agent in a simulated environment. A Q-learning algorithm is employed to simultaneously train adversarial models for both the pursuer and evader over thousands of iterations. Our simulation environment leverages the OpenAI Gym API to create a discretized gridworld space and set of actions. Sensor models provide observations to guide the agents' learning and rewards.

Introduction

Pursuit-evasion games involve one agent, the seeker or pursuer, planning movements to capture another agent, the evader, which tries to plan an optimal escape path. These scenarios have many real-world applications like search and rescue, animal herding, vehicle tracking, and assistive robotics. Developing a generalized solution to pursuit-evasion problems could enable autonomous systems to handle tasks typically requiring human intervention.

We use a Q-learning reinforcement learning (RL) algorithm to train two competing agents in the same game environment. Q-learning is a model-free RL technique that learns the value of taking actions in given states to maximize cumulative rewards over time. It is well-suited for unknown environments since it learns directly from interactions without requiring an environmental model.

Methodology

We formulate the pursuit-evasion game as a discrete environment with a 20x20 grid world and 6 possible movement actions per agent (forward, rotate left/right, move and rotate left/right, or stop). The observation space consists of the agent's position, orientation, and a binary sensor reading indicating if the other agent is visible within a 180° field of view and 5 unit range.

ACTION	LINEAR COMPONENT	ANGULAR COMPONENT
Stop	0	0
Forward	1	0
Turn Left	0	1
Turn Right	0	-1
Move and turn left	1	1
Move and turn right	1	-1

Observation	Returned
Sensed	0
Not Sensed	1

Separate Q-tables are maintained for the pursuer and evader. At each step, both agents select an epsilon-greedy action based on their current Q-values. Rewards are shaped to motivate the pursuer to capture the evader quickly, while the evader is rewarded for avoiding the pursuer.

The agents' rewards are defined as:

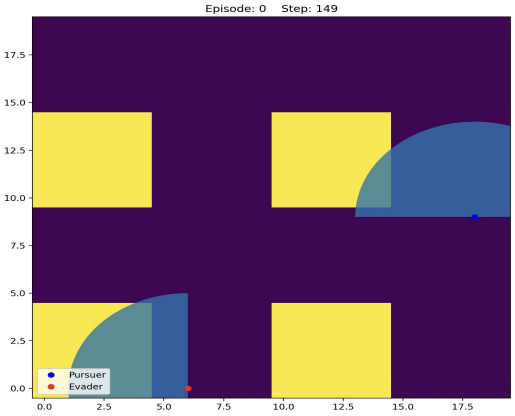
Pursuer Rewards:
- +100 for capturing evader
- +2 for observing evader
- -1 for failing to capture

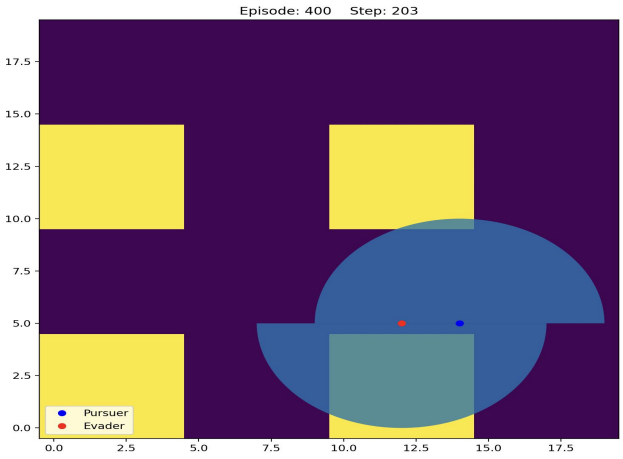
Evader Rewards:
- -100 for being captured
- -2 for being observed by pursuer
- +1 for avoiding pursuer

Using adversarially balanced rewards, the optimal pursuit and evasion policies should converge to produce equal average rewards for each agent over many episodes.

We apply the following hyperparameters: learning rate 0.01, discount factor 0.9, 200,000 training episodes with 999 maximum steps per episode.

Simulation Results





Conclusion

This project successfully applied Q-learning to produce viable adversarial pursuit and evasion policies in a simplified gridworld environment. The convergence of agents' rewards demonstrates the ability of RL to derive robust strategies from mere interaction rewards.