Justin Chevalier

CS370 Current and Emerging Trends

Southern New Hampshire University

Module Six Assignment

October 7th 2023

The reinforcement learning algorithm is a policy based technique that can be used to solve the cart pole that was featured in the last module. The goal of cart pole problem is to balance the pole on the cart by controlling the motion of the cart. The reinforcement learning algorithm is known to be a goal oriented algorithm which means that the algorithm will make selections with the intentions of gaining a reward such as solving a particular problem. The cart pole problem has two identifiable rewards that are specified. The first reward is obtain when the pole’s position is closest to the top. A reward can also be earned when the pole moves up which indicates the right direction. The agent will have to lift the pole up repeatedly away from the ground as it will continue to fall if the agent doesn’t. The reinforcement learning algorithm ultimately learns to keep the pole upward after countless training on the problem.

The Advantage Actor Critic algorithm can also be used to solve our cart pole problem. This algorithm combines two different types of reinforcement learning techniques together to create an improved extension of the policy gradient approach. The technique incorporates policy and value functions to improve the effectiveness of learning. To implement this technique into our cart pole problem we would have to initialize policy and value network with random weights, gather episodes by interacting with the environment using the current policy, estimate the increasing rewards and discover advantages for each step, computing the policy and value function gradients balancing the advantages and rewards. We also have to update the value and policy networks based their respective gradients and learning rates. This technique is repeated over a specified number of times to increase performance.

The policy gradient approach and the value based approach take different methods in how they approach the action selection. In value based approaches such as Q-Learning for example the goal of this technique is to learn a deterministic action from a distinct set of behaviors by locating the maximum value. In the policy based technique the goal is to learn a map from state to state actions which can be theoretical and works continuously in action spaces. This indicates that the policy gradient approach can solve problems that the value based approach can’t.

The actor based approach integrates both value based and policy based techniques and learn a policy as well as an estimate of the value function which allows them to manage both of them continuously in action spaces while also learning hypothetical policies while reducing the variance of the policy gradient updates. The actor critic approach is far more stable than the policy gradient method and assemble quicker than the value based approach.

Resources

Karagiannakos S. (2018 November 17).The idea behind Actor-Critics and how A2C and A3C improve them

<https://theaisummer.com/Actor_critics/>

Juliani A. (2016 December 16).Simple Reinforcement Learning with Tensorflow Part 8: Asynchronous Actor-Critic Agents (A3C)

https://medium.com/emergent-future/simple-reinforcement-learning-with-tensorflow-part-8-asynchronous-actor-critic-agents-a3c-c88f72a5e9f2