Michael Garramone

CS370 Current/Emerging Trends in CS

2/14/2023

Professor Alexander

Module 6 Assignment: Cartpole Revisited

**Explain how the cartpole problem can be solved using the REINFORCE algorithm**.

The cartpole problem is an inverted pendulum with a center of gravity above the pivot point sitting in a cart. The pendulum will move either left or right and the system is supposed to calculate the velocity at which the cart is to move to keep the pendulum upright. The REINFORCE algorithm could be used to solve the problem instead of using a DQN solution. “The REINFORCE algorithm works by using the Monte Carlo policy gradient theorem to update the policy in the direction of the maximum expected reward” (Jayakody, 2023). Through the Monte Carlo Policy gradient, parameters are updated after each step leading to a more suitable convergence to the optimal parameters (Jayakody, 2023).

**Explain how the cartpole problem can be solved using the A2C algorithm.**

**The A2C (Advantage Actor-Critic) algorithm works differently than the REINFORCE algorithm. In the A2C there is an actor and a critic. “**The actor takes as input the state and outputs the best action. It essentially controls how the agent behaves by **learning the optimal policy** (policy-based). The critic, on the other hand, **evaluates the action by computing the value function”** (value based). (Sergios Karagiannakos, 2018)**. In A2C, the critic learns the Advantage values instead of the Q values. This allows the system to learn how to better an action taking place instead of how good it is to be at a specific state (**Sergios Karagiannakos, 2018**). The advantage function reduces the high variance of policy networks and stabilizes the model.**

**Explain how policy gradient approaches differ from value-based approaches, such as Q-learning**.

“Policy gradient algorithm is a policy iteration approach where policy is directly manipulated to reach the optimal policy that maximises the expected return” (Karunakaran, 2020). Policy Gradient is a model-free reinforcement learning, which has no prior knowledge of the model of the environment. In Q-learning, there is no model of the environment that would be used as a guide for the reinforcement learning process (Kerner, 2023). Q-learning takes an off policy approach to reinforcement learning where it aims for the optimal action based on the current state. This can be accomplished either by developing a set of rules on its own or deviating from the set policy created (Kerner, 2023). Policy gradient approach differs from value-based approach in that a policy gradient approach can handle continuous actions. Value-based approach is limited to discrete actions when running.

**Explain how actor-critic approaches differ from value- and policy-based approaches**.

Actor critic approaches such as A2C use actor functions to input the state and output the best actions and the critic function to evaluate how well the actor performed or how correct the actions were for the reward. Value-based approaches work in a way that the agent learns from the cumulative reward for each state-action-reward-state-action. While Policy gradient, “…you strive to come up with a policy that allows you to get the most reward in the future by doing actions in each state (Deepchecks, 2022).

References:

Jayakody, D. (2023, September 24). *Reinforce - a quick introduction (with code)*. Dilith Jayakody. https://dilithjay.com/blog/reinforce-a-quick-introduction-with-code/

Sergios Karagiannakos. (2018, November 17). *The idea behind actor-critics and how A2c and A3C improve them*. AI Summer. https://theaisummer.com/Actor\_critics/

Karunakaran, D. (2020, June 8). *Reinforce - a policy-gradient based reinforcement learning algorithm*. Medium. https://medium.com/intro-to-artificial-intelligence/reinforce-a-policy-gradient-based-reinforcement-learning-algorithm-84bde440c816

Kerner, S. M. (2023, May 22). *What is Q-learning?*. Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/Q-learning

*What is reinforcement learning*. Deepchecks. (2022, November 20). https://deepchecks.com/glossary/reinforcement-learning/#:~:text=Reinforcement%20Learning%20Algorithms&text=Based%20on%20value%20%2DThe%20goal,of%20the%20current%20policy%20states.