Ethan Daugherty

SNHU

CS-370-H5862 Current/Emerging Trends in CS

6-2 Assignment: Cartpole Revisited

06/12/2022

**Cartpole Revisited**

A REINFORCE algorithm is a type of Policy Gradient algorithm that encourages an actor to interact with its environment based on policies created to tell the actor which actions it can take at any given state. Taking an action in the environment generates a reward based on the action outlined in the policy with a higher reward teaching the system to earn the most rewards from interacting. (Samsudin, 2020) Using this to solve the cart pole problem an Ai would reward a higher reward for balancing the pole counter in the direction to keep the center of mass above the cart.

Diagram

Description automatically generated

Figure 1 A simple REINFORCE algorithm

The actor would analyze the state of the cartpole, then by following the policy to adjust best to balance the pole, in this case by moving left the actor will work to earn the highest reward available in the environment and start the process over.

An Advantage Actor Critic or A2C algorithm is another policy gradient algorithm consisting of three components. The first neural network the actor, the second neural network the critic, and the Advantage work together, the actor to choose the action, the critic to decide the probability of the action taken and finally the Advantage predicts if the reward was better or worse than predicted and assigns value to the action taken based on this. The Advantage function determines if the new state of the pole is better or worse based on the action taken and (Wang, 2021) A A2C algorithm could be used to solve the cart pole problem by using the actor to choose an action based on a prediction assigned by the critic. After the action is take if it resulted in a more balanced pole the critic updates the prediction. The longer the pole is balanced the higher the reward assigned by the Advantage.

Diagram, schematic

Description automatically generated

Figure 2 Simple A2C used to solve cartpole

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

A policy gradient approach is a type of reinforcement learning agent focused around maximizing the reward by following a policy. A A2C algorithm is a type of policy gradient that updates the reward given if the actor chooses an action within an environment that leads to more optimal results. (Kapoor, 2018) A value-based approach updates the value of a function based on an equation. Q-Learning is a value-based reinforcement learning algorithm that updates the reward based on the action taken leading to more optimal results and feeds the updated reward into the equation to determine the next action. (Shyalika, 2019) The big difference being value-based approaches update the value of the reward based on an equation vs a Policy gradient approach that updates the reward based on actions taken within the policy generated by the Ai.

# References

Kapoor, S. (2018, June 2nd). *Policy Gradients in a Nutshell*. Retrieved from Twords Data Science: https://towardsdatascience.com/policy-gradients-in-a-nutshell-8b72f9743c5d

Samsudin, N. (2020, November 12). *medium.com*. Retrieved from REINFORCE Algorithm: Taking baby steps in reinforcement learning: https://medium.com/analytics-vidhya/reinforce-algorithm-taking-baby-steps-in-reinforcement-learning-ebb1048419e9

Shyalika, C. (2019, November 19th). *Twords Data Science*. Retrieved from A Beginners Guide to Q-Learning: https://towardsdatascience.com/a-beginners-guide-to-q-learning-c3e2a30a653c#:~:text=Q-learning%20is%20a%20values-based%20learning%20algorithm.%20Value%20based,greedy%20policy%20obtained%20from%20the%20last%20policy%20improvement.

Wang, M. (2021, Janurary 22). *Advantage Actor Critic Tutorial: minA2C*. Retrieved from Towards Data Science: https://towardsdatascience.com/advantage-actor-critic-tutorial-mina2c-7a3249962fc8

Yoon, C. (2018, December 29th). *Deriving Policy Gradients and Implementing REINFORCE*. Retrieved from Medimum.com: https://medium.com/@thechrisyoon/deriving-policy-gradients-and-implementing-reinforce-f887949bd63

Yoon, C. (2019, Feburary 5th). *Understanding Actor Critic Methods and A2C*. Retrieved from towardsdatascience.com: https://towardsdatascience.com/understanding-actor-critic-methods-931b97b6df3f