Jamie Javis

CS - 370

06/23/24

Cartpole Revisited

1. Explain how the cartpole problem can be solved using the REINFORCE algorithm. Consider using pseudocode, UML, diagrams, or flowcharts to help illustrate your solution.

The cartpole problem is an infamous idea that has been used in examples of reinforced learning. Reinforced Learning is a form of machine learning that allows AI to learn from the environment given to it. The algorithm used in the case of the Cartpole Problem was able to learn how to “win” by trial and error. The algorithm practices and will learn based off of the incentives to the “Win Causes” and “Lose Causes”.

2. Explain how the cartpole problem can be solved using the A2C algorithm. Consider using pseudocode, UML, diagrams, or flowcharts to help illustrate your solution.

The Advantage Actor-Critic Method is another reinforcement learning technique that uses an Actor (Policy method) and a Critic (Value Method). A policy is a probability of actions taken. A critique is a reward that the model is trying to achieve. The “advanced” side of this actor-critic method, applies an advantage function. The function is used to figure out the better actions that end in the best-win scenario. Using the A2C Algorithm, the cart pole problem can be solved. In this model, we would use an actor that creates a probability and then use a critic to incentivize the model to predict and choose the best actions.

Pseudocode Model:

Import tf.keras.model;

Constructor {

Int = Num\_Actions

Int = Units

Var Common, Actor, Critic }

Void Start (params) {

Var Common = ‘value’

Return Actor + Critic

Var Num\_Act

Var Num\_HidUni

Void Model(Actor + Critic)

Params Actions

Params Hidden Units

};

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

A policy-based approach incorporates a factor using a policy that would tell the factor to do actions that would maximize the reward over time. A policy gradient is a policy that allows the factor to learn the map given and become proficient. A value-based approach is a little different from the policy and policy gradients. In a value-based method, the factor learns a value function, which uses rewards based on the actions taken. The goal of the value-based approach is to find the optimal function that will give the best reward while the goal for the policy gradient is to update and learn as the model goes on. The major difference is that the Value-based model will consult a value while the policy will learn and update its value as it goes.

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

An actor-critic approach uses both an actor and a critic. The actor selects actions that are set by the policy, and the critic gives the actor a state-value function that tells the actor how successful the action was. The actor-critic approach combines the best parts of the value and policy-based approaches. The difference for the value-based approach is to learn the value function while the policy-based approach aims to learn the policy and build on it.

References

*Actor-Critic Algorithm in Reinforcement Learning*. (2023, December 13). GeeksforGeeks. <https://www.geeksforgeeks.org/actor-critic-algorithm-in-reinforcement-learning/>

*PyLessons*. (n.d.). Pylessons.com. https://pylessons.com/CartPole-reinforcement-learning

*Reinforcement Learning: Exploring Policy vs. Value-Based Methods*. (n.d.). Dataheadhunters.com. <https://dataheadhunters.com/academy/reinforcement-learning-exploring-policy-vs-value-based-methods/>

‌Yoon, C. (2019, July 17). *Understanding Actor Critic Methods*. Medium. <https://towardsdatascience.com/understanding-actor-critic-methods-931b97b6df3f>

*What is the relation between Q-learning and policy gradients methods?* (n.d.). Artificial Intelligence Stack Exchange. https://ai.stackexchange.com/questions/6196/what-is-the-relation-between-q-learning-and-policy-gradients-methods

‌

‌

‌

‌