Reinforcement learning runs through the algorithm multiple times and learns from rewards and punishments. The cartpole problem does this exact thing. It is able to run through and learn what choices are best made to solve the problem. The more it can run through the problem, the more information it is able to learn. Therefore, the more the problem is run, the more accurately it can make correct choices. With the rate of correct choices, it can start solving the problem at a faster rate. This will allow the problem to be solved in a more proficient way.

The A2C algorithm is a type of reinforcement learning. It utilizes both policy-based and value-based methods. Each one is able to work against the limitations of each of the other when combined. (GeeksforGeeks, 2024) This can be useful for the cartpole problem. It would be able to use the policy-based side to help make decisions based on what it already knows, while the value-based side is able to validate the rewards of such a decision. It will be able to use this to know at a fast pace if a decision is made that would be beneficial. If the outcome looks good, the agent is more likely to make that decision and move forward. This also in turn allows the agent to make decisions a bit more quickly.

Gradient approaches measure and predict possible errors. With this information, the agent processes, so it is possible to change course in a decision to achieve a better-predicted outcome. A value-based approach looks for potential rewards. In this case, it will make whatever decision it can to achieve better rewards over the least amount of error. Due to their own limitations, this method is usually combined with other methods to improve the outcome.

Actor-critic approaches work by combining both value-based and policy-based approaches. Individually, each approach has its relevance. Value-based approaches will allow it to make decisions based on what will have the better reward. In contrast, policy-based approaches will allow it to utilize already known and learned information. When combined, these two approaches work together to help cover each limitation of each. This, on the other hand, allows for more accurate and faster learning. Therefore, the actor-critic method becomes a better-suited approach for multiple different problems. (ScienceDirect, *Gradient method*)

**References**

GeeksforGeeks. (2024b, March 22). *Actor-critic algorithm in Reinforcement Learning*. https://www.geeksforgeeks.org/actor-critic-algorithm-in-reinforcement-learning/

*Gradient method*. Gradient Method - an overview | ScienceDirect Topics. (n.d.). https://www.sciencedirect.com/topics/computer-science/gradient-method

Juliani, A. (2017, June 30). *Simple reinforcement learning with tensorflow part 8: Asynchronous actor-critic agents (A3C)*. Medium. https://awjuliani.medium.com/simple-reinforcement-learning-with-tensorflow-part-8-asynchronous-actor-critic-agents-a3c-c88f72a5e9f2

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

Surma, G. (2019, November 10). *Cartpole - introduction to reinforcement learning (DQN - deep Q-learning)*. Medium. https://gsurma.medium.com/cartpole-introduction-to-reinforcement-learning-ed0eb5b58288

Yoon, C. (2019a, May 23). *Deriving policy gradients and implementing reinforce*. Medium. https://medium.com/@thechrisyoon/deriving-policy-gradients-and-implementing-reinforce-f887949bd63

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