Benjamin Houldridge

Current/Emerging Trends in Computer Science

CS 370 11999-M01

Module Six Assignment

A policy-based approach uses the model to determine action that should be taken given the current state of the game. When training a model this way, we have it play games and save the decisions and outcomes. Then using an algorithm, we determine what tweaks need to be made to each weight in the model to produce a better outcome. This algorithm produces a gradient or uses the gradient. I’ll be honest the math behind the REINFORCE algorithm went right over my head. The REINFORCE algorithm is a way to determine the gradient of the policy so that it can be updated to produce higher returns. This is an important step in the training of a model using a policy-based approach.

An actor-critic approach is different from the policy approach above. In this approach we use two networks. The actor determines what action to take, more right or left and the critic determines what the long-term score is likely to be. We then train the two models off each other. Using the actor to train the critic and the critic to score the actor. Each gets updated until stability is reached.

Policy gradient, value-based, and actor-critic approaches all differ and have their benefits and drawbacks. Policy gradient approaches learn the policy directly. They work well with continuous action spaces like the cartpole problem but are often inefficient. Value based methods like Q-learning learn to estimate the expected return of actions to make decisions. They are often more efficient to train but don’t do as well with continuous action spaces like with the cartpole problem. Action-critic methods use two models, one to learn the policy and one to learn the value function. This can make them even more efficient but with the possibility that the models never converge. This method is also more complex to implement and tune.

Sources

Yoon, Chris. “Deriving Policy Gradients and Implementing REINFORCE.” *Medium*, 23 May 2019, medium.com/@thechrisyoon/deriving-policy-gradients-and-implementing-reinforce-f887949bd63.