RFP3

Zach Majernik, Reid Arendale, Logan Kerns

What Qualifies “The Best”?

We as a team have decided on two metrics for deciding what defines the “Best” algorithm:

- How long it takes to converge
- How often it converges on the optimal path

Since the actual time in seconds it takes to converge can depend entirely on the render_mode of the maze, We used the number of episodes it takes to converge to measure time as it is more reliable.

We also use how often it finds the best path as a metric because an algorithm that converges in 100 episodes on average but finds the best path 90% of the time is better than an algorithm that converges a few episodes faster but only converges 50% of the time.

What Rewards to Use?

We ran two sets of trials for each algorithm, each using a different set of rewards.

Reward set 0:

- +10 reward for reaching the present
- Like a “control” test

Reward set 1:

- +10 reward for reaching the present
- Reward equal to ½ change in distance towards present
  - Negative if its away, positive if its towards
- Negative reward for revisiting spaces equal to -0.25 x timesVisited
  - If timesVisited is 0, reward is 0, but if times visited is 5, reward is -1.25

Q-Learning

SARSA

Monte Carlo - **First** Visit

Monte Carlo - **Every** Visit

And the Winner is...