Q2: Solving Search Problems with MDPs

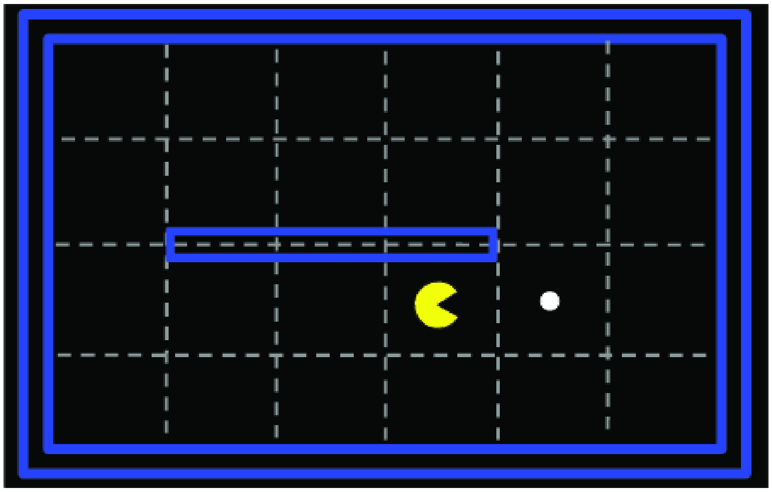
Problem 2: Solving Search Problems with MDPs

The following parts consider a Pacman agent in a deterministic environment. A sink state is reached when there are no remaining food pellets on the board, which terminates the MDP. Pacman's available actions are $\{N, S, E, W\}$, but Pacman **can not** move into a wall. Whenever Pacman eats a food pellet he receives a reward of +1.

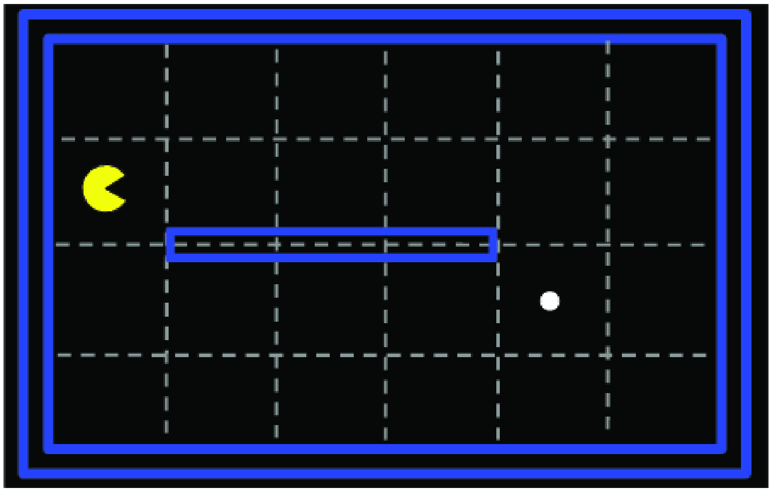
Assume that pacman eats a food pellet as soon as he occupies the location of the food pellet-i.e., the reward is received for the transition into the square with the food pellet.

Consider the particular Pacman board states shown below. Throughout this problem assume that $V_0(s) = 0$ for all states, s. Let the discount factor, $\gamma = 1$.

State A:



State B:



Part 1

0.0/3.0 points (ungraded)

What is the optimal value of state A, $V^*(A)$?

1

Answer: 1

Submit

You have used 0 of 2 attempts

Answers are displayed within the problem

Part 2

0.0/3.0 points (ungraded)

What is the optimal value of state B, $V^*(B)$?

1

Answer: 1

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Part 3

0.0/3.0 points (ungraded)

At what iteration, k, will $V_k(B)$ first be non-zero?

5

Answer: 5

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Part 4

0.0/3.0 points (ungraded)

How do the optimal q-state values of moving W and E from state A compare?

- ( ) $Q^*(A, W) < Q^*(A, E)$
- ( ) $Q^*(A, W) > Q^*(A, E)$
- (•) $Q^*(A, W) = Q^*(A, E)$ ✔

Submit You have used 0 of 1 attempt

Answers are displayed within the problem

Part 5

0.0/6.0 points (ungraded)

If we use this MDP formulation, is the policy found guaranteed to produce the shortest path from pacman's starting position to the food pellet? If not, how could you modify the MDP formulation to guarantee that the optimal policy found will produce the shortest path from pacman's starting position to the food pellet? Select all answers that apply, there may be more than one.

- [ ] Yes, the formulation is guaranteed to produce the shortest path.
- [x] No, the formulation is not guaranteed to produce the shortest path. Adding a negative living reward will guarantee a policy producing the shortest path. ✔
- [ ] No, the formulation is not guaranteed to produce the shortest path. Adding a positive living reward will guarantee a policy producing the shortest path.
- [x] No, the formulation is not guaranteed to produce the shortest path. Setting the discount factor to a value between zero and one will guarantee a policy producing the shortest path. ✔
- [ ] No, the formulation is not guaranteed to produce the shortest path. Setting the discount factor to a value greater than one will guarantee a policy producing the shortest path.
- [ ] No, the formulation is not guaranteed to produce the shortest path. Changing the reward function so that it returns -1 for any action that would cause pacman to move into a wall will guarantee a policy producing the shortest path.
- [ ] No, the formulation is not guaranteed to produce the shortest path. It is not possible to modify the MDP such that it guarantees a policy that produces the shortest path.

Submit

You have used 0 of 2 attempts

Answers are displayed within the problem