2. Home World Game

Project due May 10, 2023 08:59 -03 Completed

In this project, we will consider a text-based game represented by the tuple $< H, C,$ the set of all possible game states. The actions taken by the player are multi-word natu such as **eat apple** or **go east** . In this project we limit ourselves to consider commands (e.g., **eat**) and one argument object (e.g. **apple**).

$C = \{(a, b)\}$ is the set of all commands (action-object pairs).

$P : H \times C \times H \to [0, 1]$ is the transition matrix: $P\left(h' | h, a, b\right)$ is the probability of r command $c = (a, b)$ is taken in state h.

$R : H \times C \to \mathbb{R}$ is the deterministic reward function: $R\left(h, a, b\right)$ is the immediate rev when taking command (a, b) in state h. We consider discounted accumulated rewards factor. In particular, the game state h is **hidden** from the player, who only receives a va

Let S denote the space of all possible text descriptions. The text descriptions s observ produced by a stochastic function $\Psi : H \to S$. Assume that each observable state s **unique** hidden state, denoted by $h\left(s\right) \in H$.

You will conduct experiments on a small Home World, which mimic the environment of world consists of four rooms- a living room, a bed room, a kitchen and a garden with co (illustrated in figure below). Transitions between the rooms are **deterministic**. Each roo representative object that the player can interact with. For instance, the living room has can **watch** , and the kitchen has an **apple** that the player can **eat.** Each room has sever randomly on each visit by the player.

Rooms and objects in the Home world with connecting pathwa

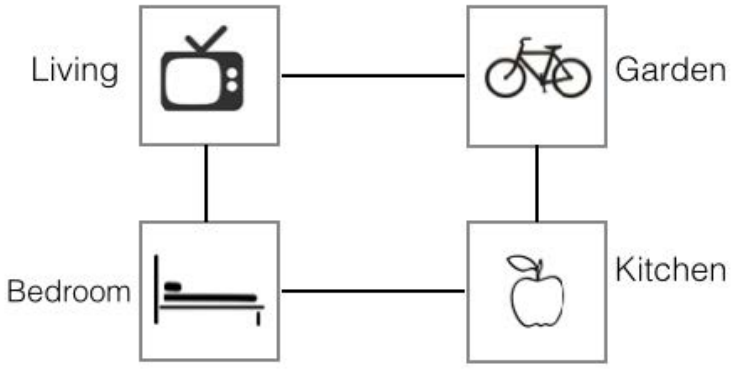

Table 1: Reward Structure

Positive	Negative
Quest goal: $+1$	Negative per step: -0.01
	Invalid command: -0.1

(so as the quest description). When the player finishes the que
after time are assumed to be zero, i.e., for . Over the course of the ep
discounted reward obtained by the player is

We emphasize that the hidden state are unobservable to the player.

The learning goal of the player is to find a policy that that maximizes the ex
discounted reward where the expectation a
randomness in the model and the player. Let denote the optimal policy. For each ob
let be the associated hidden state. The optimal expected reward achievable is de

where

We can define the optimal Q-function as

Note that given , we can obtain an optimal policy:

The commands set contain all pairs. Note that some commands ar
(eat,TV) is invalid for any state, and **(eat, apple)** is valid only when the player is in the k
corresponds to the index of kitchen). When an invalid command is taken, the system st
and a negative reward is incurred. Recall that there are **four** rooms in this game. Assum
quests in this game, each of which would be finished only if the player takes a particula

? STANDARD NOTATION

sum_t(gamma^t*R_t)

Submit You have used 1 of 6 attempts

Relation between value function and Q-function

1/1 point (graded)

Which of the following equation gives the correct relation between and ?

Submit You have used 1 of 4 attempts

Optimal episodic reward

1/1 point (graded)

Assume that the reward function is given in Table 1. At the beginning of eac player is placed in a random room and provided with a randomly selected quest. Let value function for an initial state , i.e.,

Home World Game

Show all posts

? a little confused about the last question
I'm sorry if my question seems dumb, but are we expected to write a code that will calculate the answer, or t

? [Staff] receiving error message
I see "Could not format HTML for problem. Contact course staff in the discussion forum for assistance." whe

? clarification on mechanism to collect rewards
In the text is written that I receive a small negative reward for every non-terminating step, does this mean th

Read TAB 4 first - it has a better description of the game than this confusing mess
Read this section: Evaluating Tabular Q-learning on Home World

Silly question but what's the ~ mean in this context?

? [Staff] I entered 6 incorrect answers but it doesn't have "Show Answer"
Hi, I entered 6 incorrect answers on the third question "Optimal episodic reward" but it doesn't have "Show A

Optimal Episodic reward
It says the episode ends either when the model finishes the quest or exceeds the maximum number of steps

? Optimal episodic reward - Confused
Not getting the right answer on optimal episodic reward. Let me show my reasoning. In my opinion there are

? Question about standard notation

? the format of the answer in the last question
Dear all, Can anyone explain the format of the results in the last question? There are 4 by 4 states, we can ca

The states are hidden
Can someone explain to me what it means when it says the states are hidden? Is the model capable of memo

? Determining Episodic Reward without knowing the path

Clarifying Table 1

Relation between value function and Q-function
I am lost in this particular manipulation of the equations. Even the form of the answers where Q is a function

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap
Cookie Policy
Do Not Sell My Personal Information

Connect

Blog
Contact Us
Help Center
Security
Media Kit