CSCE 475/875
Game Day 1: Learning Day
Assigned: February 10, 2020 Game Day: February 18, 2020

Introduction

Welcome to Soh's Agency for Multiagent Learning. Today, we have a set of states and a set of actions for you to try out in order to learn about the Q-values of the state-action pairs. The objective is to learn about Reinforcement Learning in a multiagent setting, insights as to how to set the learning rate and discount factor, and how to compute the Q-values and values of states.

This document is supplementary to Game Day 1: Learning Day assignment.

Virtual Currency

As you perform actions online, you will be given virtual rewards and those count towards your virtual currency. You may use the virtual currency to pay for at most 2 matrices ($2000 for the first one, and $1000 for the second one) after Round 1. If you do not have enough virtual currency at the time of transactions, that would still be fine as the amount will be deducted from your total at the end of Round 2.

Procedure

1. Please go to the following URL: TBA at game time.
2. Select your team and click OK.
3. After logging in, you should be at a page that asks you to "Start Here". Click on the link.
4. Now you are on the participation page.
5. *Prior to this, you should have already decided on your α and β values and have included them in your pre-game strategy that you should have turned in by now.*
6. When the Game Day Monitor announces the start of Round 1, please select the action that you would like to perform, observe the reward and the resulting state, and keep track of your observations to help you compute your Q-values.
7. When the Game Day Monitor announces the end of Round 1, then you must stop submitting actions to the participation page. You must then immediately e-mail your Q-value matrix to the Game Day Monitor within 3 minutes of the end of Round 1.
8. During the intermission, you will be allowed to approach other teams to find out more about their Q-value matrices. However, there must be at least one person at your station so that when the other teams approach you, you will be able to provide information as you wish. *Note that each team is not required to be truthful or rational when gathering or providing information about their Q-value matrices.*
9. During the intermission, please also decide whether to purchase matrices from the Game Day Monitor. To do so, e-mail the Game Day Monitor with your request clearly. Each team

is allowed to only make two purchases: $2000 for the first matrix, and $1000 for the second. 50% of each purchase will go to the team whose matrix is purchased.

10. If you replace your own Q-matrix with any of the two matrices that you have purchased, please clearly indicate so on the worksheets.
11. Prior to the start of Round 2, *you must again determine your α and β values. Please note them and justify your choice on the worksheets.*
12. Then Round 2 begins. Proceed following the same set of activities as you have done in Round 2.
13. At the end of Round 2, please submit your final Q-value matrix to the Game Day Monitor via e-mail.
14. Then, please complete your post-game analysis.

Other Logistics

Round 1: 15-25 minutes

Intermission: 5-15 minutes

Round 2: 15-25 minutes

Post-Game Analysis: 5-10 minutes

You are required to keep track of all your transactions in this worksheet. Note that ideally, after each transaction, you should compute and update the Q value of the state-action pair and the value (V) of the state involved. *You may want to implement this program before the Game Day and use the program to help you do this accurately and efficiently during Game Day*.

Round 1

Learning Rate: _____ Discount Factor: _____

Trans.	State	Action	Rewards		Action By	Q(state,action)	V(state)
			States	$			
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							

Trans.	State	Action	Rewards		Action By	Q(state,action)	V(state)
			States	$			
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							
61							
62							
63							
64							
65							
66							
67							
68							
69							
70							
71							
72							
73							
71							
72							
73							
74							
75							

Intermission

During **Intermission**, each team will e-mail the Monitor their ordering of their state-action pairs. And then the Monitor will share them with all teams. Also, after **Round 2**, each team is also required to e-mail the Monitor their ordering of their state-action pairs.

Let us do this in Excel. Please name your file in this manner: Round1_<Team Name>.xlsx
So, if your team name is "Foobar", then: Round1_Foobar.xlsx

In the file, please have the following (just an example, the Q values are just some random numbers). Since there are only 6 states x 3 actions, there are only 18 state-action pairs. The state-action pair that has the highest Q value is given the order 1, the second highest the order 2, and so forth. With this format, it will allow us the monitors to quickly merge all orderings into one and send it back to you.

	<TEAM NAME>	
Pair	Q	Order
S1, A1	102	6
S1, A2	129.21	3
S1, A3	20	12
S2, A1	30.1	11
S2, A2	40.2	9
S2, A3	60.19	8
S3, A1	394	1
S3, A2	102	5
S3, A3	8.2	13
S4, A1	33.11	10
S4, A2	203.12	2
S4, A3	94	7
S5, A1	109	4
S5, A2	1	14
S5, A3	0.5	15
S6, A1	0.5	15
S6, A2	0.5	15
S6, A3	0.5	15

Notes:

Information Gathered:

Information Provided:

Purchase 1: Q-value matrix (Which team) _____ Amount Paid _____

Purchase 2: Q-value matrix (Which team) _____ Amount Paid _____

Replace own Q-value matrix? If yes, How?

Round 2

Learning Rate: _____ Discount Factor: _____

Trans.	State	Action	Rewards		Action By	Q(state,action)	V(state)
			States	$			
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							

Trans.	State	Action	Rewards		Action By	Q(state,action)	V(state)
			States	$			
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							
61							
62							
63							
64							
65							
66							
67							
68							
69							
70							
71							
72							
73							
71							
72							
73							
74							
75							

POST-GAME ANALYSIS

Notes: ***Lessons learned, reflections, etc.***

AMMOUNT OF REWARDS ($): _____

SALES ($): _____

PURCHASES ($): - _____

GRAND TOTAL: _____