

AI

ASSIGNMENT-2

Yash Patel
Team : 75
201301134
CSE, IIIT-H

PART-A



Problem Details :

- 1). $X = 75$ (team numebr = 75)
- 2). $\text{Gamma} = 1$
- 3). $\text{Delta} = 3.5$
- 4). Actions = {'move_north', 'move_south', 'move_east', 'move_west'}
- 5). $P(\text{intended_action}) = 0.8$; $P(\text{perpendicular_action}) = 0.1$ (each).
- 6). $R(s,a) = -3.5$ for non-terminal states.
- 6). No action is performed at terminal states.

Problem Statement :

Part A : Perform the Value Iteration algorithm manually on the above MDP to calculate the reward achieved for the given start state. The cell (0,1) is the positive sink whereas cell (1,0) is the negative sink. The dark colored cell is blocked(assume it is a wall). All the four corner sides of the matrix are also considered to be walls. Replace X with your respective reward value. The parameters gamma and delta are as mentioned above.

Team number = 75;

ITERATION: 0.000000

48.750000	75.000000	56.250000
-75.000000	48.750000	-3.750000
-3.750000	-3.750000	-3.750000
-3.750000	-3.750000	0.000000

ITERATION: 1.000000

53.625000	75.000000	61.500000
-75.000000	48.375000	45.750000
-7.500000	34.500000	-7.500000
-7.500000	-7.500000	0.000000

ITERATION: 2.000000

54.112499	75.000000	66.974998
-75.000000	53.325001	54.862499
15.600000	33.450001	35.549999
-11.250000	22.350000	0.000000

ITERATION: 3.000000

54.161251	75.000000	68.433746
-75.000000	54.236252	60.648750
14.385000	44.025002	47.040001
14.565001	24.120001	0.000000

ITERATION: 4.000000

54.166126	75.000000	69.158249
-75.000000	56.671501	62.485497
25.426502	45.781502	53.875500
18.441000	35.338501	0.000000

ITERATION: 5.000000

54.166611	75.000000	69.414375
-75.000000	58.316547	63.492302
27.219301	49.517403	56.204098
28.907551	38.253151	0.000000

ITERATION: 6.000000

54.166660	75.000000	69.540665
-75.000000	59.495583	63.962387
31.254677	51.245579	57.615990
32.465206	42.579994	0.000000

ITERATION: 7.000000

54.166664	75.000000	69.600304
-75.000000	60.044468	64.228333
32.992985	52.733532	58.306068
36.685982	44.750984	0.000000

ITERATION: 8.000000

54.166668	75.000000	69.632866
-----------	-----------	-----------

-75.000000 60.406021 64.357521
34.605423 53.415482 58.736626
39.018684 46.580521 0.000000

ITERATION: 9.000000

54.166668 75.000000 69.649040
-75.000000 60.577568 64.432648
36.267036 53.937954 58.951229
40.876827 47.542305 0.000000

ITERATION: 10.000000

54.166668 75.000000 69.658173
-75.000000 60.689911 64.470253
37.971962 54.233879 59.085037
41.998230 48.242275 0.000000

ITERATION: 11.000000

54.166668 75.000000 69.662842
-75.000000 60.749588 64.492554
39.069168 54.507629 59.158092
42.840839 48.661156 0.000000

ITERATION: 12.000000

54.166668 75.000000 69.665543
-75.000000 60.794807 64.504486

39.880352 54.672398 59.210617
43.369926 49.006302 0.000000

ITERATION: 13.000000

54.166668 75.000000 69.667000
-75.000000 60.820831 64.512360
40.401215 54.794945 59.241890
43.780067 49.225540 0.000000

ITERATION: 14.000000

54.166668 75.000000 69.667938
-75.000000 60.839386 64.516922
40.793671 54.870975 59.263573
44.048561 49.386517 0.000000

Policy according to MDP (MAPPED TO GRID):

RIGHT	TERMINAL	LEFT
TERMINAL	RIGHT	UP
RIGHT	UP	UP
LEFT	UP	NONE

Path from start:

LEFT->UP->UP->RIGHT->UP->LEFT

PART-B

Part B : Modelling the above problem using LP show below.

Q1: Model the parameters r , A and α

Q2: Use the excel LP solver to compute the x values and the expected reward for this

MDP

Q3: Please verify that the expected reward obtained is equivalent to the one obtained using the VI algorithm. The VI value and LP value can differ at max by $\Delta * 1.2$.

Answer : The expected reward from LP is 38.68715929 which is quite close to Value Iteration which is 39.018684(after 8th iteration).

Other parameters are mentioned in excel.