## MDL-Assigmment-3

ROMICA RAISINGHINN Implementing value Steration for MDP 2021101053

According to value steration rigosimm,

we initiatize the utilities of each state to zero.

i.e, uo(I)=0

· Uo(I) > willy of state I at time

In the question given: utility of goal state = +; Utility of penalty state = - 1 step cost = -0.04Discount factor = 8 = 0.95

sterate: Step 2:

PROB of going According to Bellman Update Equation: in the disrection ( of action

Ut+1(I)=max[R(I,A)+8 \( P(I)I,A) \tag{1}

where,

R(I, A) - Dimmediate reward

Reward on taking action A on state I (step cost)

P(I/I,A) - D transition probability on reaching state I from state I upon

U+(I) -> Utility value of state I in me previous îterali on

y > discount factor

we vous continue iterating until the values [Ut, Ut+1] converge.

At the end of iteration, we calculate optimal policy:

Policy (I) = argma x[R(I,A)+8 \ P(I)II,A) x Ut(I)

Prob of going perpendicular 50 direction of action

prob of going opposite to the distection of action

77 (0,0) Grow

```
According to againsm:
   Initial utilities of all states
                 = u[i][j] =0
               fox ,
 For (0,0):
  -0.04 + 0.95 [0.7x0+0.150+0.160] -0.04+ 0.95 [0.7x1+0.180
  North-D
         = -0.04
 South -D
  -0.04+0.95 TO-7KO+0-15KO+7
                       D. ISXD J
          = -0.04
West/East -D
   -0.04 + 0.95 [07x0+0.15x0+
                      0-15 x0)
  mar of all actions = -0.04
           4[0][0]=-0.04
For (0,1): Groat state
          U[0][1] = 1 -
 For (0,2): Penalty state
             u[0][2] = -1
 For (1,0):
North-D
   -0.04 + 0.95[0.7x0+0.15 x0+ (1/+1) 3.07=
      = -0.04
 South -
  -0.04+0.95[0.7x0+0.15/0+
           = -0.04
West/East -D
```

= - 0.04

max of all directions = -0.04 40-0- E07CIJ U FOR (111): North-D = -0.04 + 0.665 = 0.625 South -D -0.04 +0.95 [0.7×0+ 0.15x0+0.15x0 = -0.04 West/East-D -0.04+ 0.95 [0.7 XO + 0.5X] +0.15x01  $=-0.04 + 0.95 \times 0.15$ = 0.1025 => |u[i][i] = 0.625] FOR (112): Morth: LEITO -0.04+0.95[0.7×-1 0.15×0) 0.15x0] = -0.04 -0.95x0.7 = - 0.705 south: -0.04+ 0.95[0.7x0+ 0.15x0 HO -+ 0.15x0) 10.15 x 00 (00) + 400 2 2 1000 4 + 100.00 - 1000 x west/East: -0.04+0.95[0.7x0+0.15x0+0.15x0)-0.04+0.95[0.7x0+0.15x(-1) - 40-15×0-

```
=-0.04-0.95×0.15
    -0.1825
at of all directions = -0.04
       u[1,2] = -0.04
ENS (210):
-0.04 +0.95 [0.7x 0+0.15x0
                 +0.15x0]
  = -0.04
0.04 to.95 [0.7 to t 0.15 kot
    = -6.64=160111
70x210+0x5.0] 20.0+ NO.0
   = 20.04 DJ20.0 ANO.0-1
nar of au diréctions = -0.04
      JU[2][0] = -0.04
For (2/2): Wall - 110-0
     u[2)[2] =0
For (2/2) 31 10 1200 1200
-0.04+0.95[0.7×0+0.15×0+
       = -0.04
-0.04 + 0.95[0.7x0 + 0.15x0+
 south -D
 West/East -> 1/20.00 0.15x0+
                   DISKOJ
```

= -0.04

max of all directions = -0.04 u[2][2] = -0.04] For (3,0): -0.04 +0.95 [0-7 × 0 + 0.15 × 0 North -> (0,=,-0.04") -0.04+0.95[0.7x0+0.15x0+ South -D (uhilist -0.04 majee. 0 1100 -0.04 + 0.95 [0.7x0+ 0.15x0 East/west-D = -0.04 max of an directions = -0-04 u[3][0]=-0.04 For (3,1): -0.04+ 0.95 [0.7x0+ 0.15x0 = -0.04 -0.04+0.95 [0.7x0+0.15x0] 1210 = -0.04 West/East -0 -0.04+0.95[0.7x0+0.15x0] = -0.04 max of all dis = -0.04 14 [3][1] = -0-04

-0.04+0.65[0.7×0+0.15×0+0.15×0] = -0.04

-0.04+0.95[0.7x0+0.15x0+ 0.15x0) = -0.04

West/ East -D -0.04 + 0.95 [0.7x0+0.15x0) = -0.04

max of an dishs = -0.04

u[3][2] = -0.04

## ITERATION 2:

For (0,0):

North D +0x 10/20 -0.04+0.95[0.710+0.1510] 40.15 ×1

= -0.04 +0.95×0.15 = 0.1025

-0.04+0.95[0.7×[-0.04]+6.15×0

= x004 - 0.0666

West -> File Way 12 AND -0.04 + 0.95 [0.7x0+ 0.15x0 + 0.15(-0.04)

= -0.0457

-0.04+0.95[0.7×1+0.15×0 = -0.04 + 0.95 [0-7-0.000] = 0.6193 max of all directions = 0.6193 n[0](0)=0.6133

FOR (0,1): Goal State u(0)[1) = )

For (0,2): Penalty State u[0][2] = -1

For (1,0):

Nosth ->

-0.04+0.95[0.7[-0.04] 2x310011 0.15 x 0.625]

=-0.04+0.95[-0.028

= -0.04 + 0.95 [0.6575] = 0.0224625

South -D

2010

-0.04+0.95[0.7(-0.04) +0.75×0 +D.15×0.625

- 0.0224625

-0:04+0.95[0-7X0+0.15(-0.04) +0.15 [-0.04)

= -0.04-0.0114 = 0.0514 = Jacol 1004

```
0.04+0.95[0.7×0.625+
                              max of all directions =
                                    u[1](1) = 0.62386
         0.15(-0.04)+
            0.151-0.04)
, 0.04 + 0.95 [0.4375. - U.012]
                              For (1,2):
max of all directions = 0.364225 _0.04+0.95[0-7[-])
                                            + 0.15×0
                                            + 0.15 (0.625)]
     u [1)[0] = 0.364225
                                    = -0.6159375
                              -0.04+0.95[0.7[-0.04]+0.1560
tex (11):
                   5- 1211)
North - D | 0.15 (-0.04)

North + 0.95 [0.7 × 1 + 0.15 (-0.04)]
                                    = 0.0224625
                              West $ 0.95 [0.7 (0.625) +
     = 0.62386
                                            70:15.(-0.04)
      Max of all distillent
                                    = 0.227425
    0 - - Lol(E) N
-0.04+ 0.95[0.7×0+ 0.15l-0.04]
                               [UT (40.0.) [.0] 28.0 TUS
                               East -D
                              -0.04 +0.95 [0.7x0+ 0.15 (-D)
                                               + 0.15 (-0.04)]
          0.0514 tatron
  - 12/0 00 0 × 1.0 /20.0 + 40.5
                            (41.0-)21.0+0.1887 20.0+10.0
                            max of all directions = 0.227425
  (PO.0-121.0+
West -D (20.04) +
-0.04 + 0.95 [0.7 (20.04) + 0.15x0
                                  U[1][2]=0.227425
 E, 0,0759 0 + 40.0
                               FOR (2,0) :
      Piedn
                                [0+. (40.0-) 5.0] 20.0+ 40.0-
                                North -D
-0.04 + 0.95[0.7[-0.04] +0.15x0]
                                  = -0.0666
                                                   d-intro 4
                                South - 640.0-) 1-0/20.0
     = 0,0759
                                -0.04+0.95[0.7x(-0.01)+0]
  O- was in the formation
                                   - -0.0666
 4720.0- = EIJ[E]
```

```
West/East -D
 -0.04+0.95[0.7×0+0.15(-0.04)
                 40.151-0-04)
    = -0.0514
max of all directions = -0.0514
       u[2][0] = -0.0514
 por (2,1): wall
        u[2](1) = 0
for (212):
 North 20
-0.04 + 0.95 [0.7x (-0.04) +0]
   = -0.0666
 South -D
-D.04+0.95[0-71-0.04) +0]
        - - 0.066L
 West/East >>
 -0.04+0.95[0.7x0+0.15(-0.04)
                 10.15(-0.04)
   10.0514
  max of au directions = -0.0514
         U[2][2] = -0.0514
  For (3,0):
   NOAH-D
  +0+(va.o-) F-0]20,0+va.a-
                0.15(-0.04)]
       = -0.0723
```

South-D -0.04+0.95[0.7x0 +0.1540 (40.05 l-0.04) - 6.0457 west -D -0.04 + 0.95 [0.7x0 +0.1540 +0.151-0.097 = -0.0457 Fast -D -0.04+0.95[0-7(-0.04)+ + (40.0-121.0) 0.15,0) = -0.0723 max of all directions = -0.045] u [3)[0] = -0.0457 For (3,1): -0.04+0.55[0.7×0+0.15(-0.04) +0.15(-0.04)] t=10-0.0514 -0.04 + 0.95 [0-7×0 + 0.15 (-0.04) South -D = -0.0514 West/East -D. -0.04+0.95[0.7[-0.04) +0+0] = -0.0666 max of all directions = -0.0514 u [3][1] = -0.0514

```
For (3,2):
10.04+0.95[0.7(-0.04) + 0.15(-0.04)
North -D
      _ -0. 0723
(40.04 +0.95 [0.7x0 +0.15 (-0.04)
    = -0.0457
10.04 + 0.95 [ 0.7 (-0.04) +
            0.151-0.04)+0]
       FR40,00 =
        =-0.0723
-0.04 + 0.95 [0.7x0 +0.15 (-0.04)
       = -0.0457
max of all direction= = 0.0457
     u [3][2] = -0.0457
 Matrix after Iteration 1:
                    - 1
             1
 0.62499
            0-6249.. -0.64
  -0.04
                     -0.04
               0
  -0.04
                    -0.04
           -0.04
 -0.04
 Matrix after steration 2:
                       -1
              1
 0.70836
                   0.2274
  0.4589 0.6135
                    - 0.0779
               0
  -0.0779
          -0.0779 -0.0779
  -0.0779
```

Both iteration values marches noith the code in python.

tilque, the algorithm works by manual calculation as well as on code.

There are 22 iterations taken for the algorithm to converge.