INDIANA UNIVERISTY BLOOMINGTON

PA 3: Reinforcement Learning

I 526 Applied Machine Learning

SHORT REPORT

Submitted by

Nihar Khetan
Masters Candidate in Computer Science
School of Informatics and Computing

Indiana University
nkhetan@indiana.edu

Ghanshyam Malu
Masters Candidate in Computer Science
School of Informatics and Computing
Indiana University
gmalu@indiana.edu

Under the Guidance of:

Professor Sriraam Natarajan
Asst. Professor, School of Informatics and Computing
Indiana University

Contents

Usage

```
@author   : Nihar Khetan, Ghanshyam Malu, Xiao Liang
@desc     : Grid world for reinforcement learning
            1. Report learned values by value iteration
            2. Implement Q learning with initial e = 0.9
            3. Set reward at each step to be 0. Report results.

@Usage    : Execute the python file "RUN_ME.py" to run the Gold Explorer
              $ python RUN_ME.py

@Version  : Uses Python 2.7
```

Value Iteration

With Reward -1

```
 ==============================================================================
                  Welcome to Gold Explorer Using Reinforcement Learning
 ==============================================================================

Choose one of the available options:
 0 - Explore Gold using Reinforcement Learning - Value Iteration
 1 - Explore Gold using Reinforcement Learning - Q Value

Your choice from [0, 1]... 0
Show detailed log (Y/N)?... : n
Set reward for each block preferred option [0 or -1]... :   -1
 ============================================================================
       Welcome to Gold Explorer Using Reinforcement Learning - Value Iteration
 ============================================================================

 Iterating..........

%%%%%%%%%%%%%%%%%% Total # of Value Iterations :271 %%%%%%%%%%%%%%%%%%

               *************************************************
                           Grid World Reward Matrix
               *************************************************
                 |    -1    |    -1    |    -1    |    10    |
                 ---------------------------------------------
                 |    -1    |   -50    |    -1    |    -1    |
                 ---------------------------------------------
                 |    -1    |    -1    |    -1    |    -1    |
                 ---------------------------------------------
                 |    -1    |     0    |    -1    |     0    |
                 ---------------------------------------------
                 |    -1    |    -1    |    -1    |    -1    |
                 ---------------------------------------------


    ***********************************************************************
                          Grid World Value Matrix
    ***********************************************************************
       |   36.067672   |   43.870462   |   63.667549   |   76.643424   |
       -----------------------------------------------------------------
       |   30.449663   |   -5.389848   |   52.690462   |   63.667549   |
       -----------------------------------------------------------------
       |   30.254205   |   37.072329   |   43.610152   |   52.690462   |
       -----------------------------------------------------------------
       |   25.345155   |    0.000000   |   37.072329   |    0.000000   |
       -----------------------------------------------------------------
       |   21.034770   |   25.345155   |   30.254205   |   26.228784   |
       -----------------------------------------------------------------

 ==============================================================================
   Thank you for using Gold Explorer Reinforcement Learning - Value Iteration
 ==============================================================================
```

With Reward 0

```
*************************************************
            Grid World Reward Matrix
*************************************************
  |     0     |     0     |     0     |    10     |
  ---------------------------------------------
  |     0     |    -50    |     0     |     0     |
  ---------------------------------------------
  |     0     |     0     |     0     |     0     |
  ---------------------------------------------
  |     0     |     0     |     0     |     0     |
  ---------------------------------------------
  |     0     |     0     |     0     |     0     |
  ---------------------------------------------
```

```
*********************************************************************
                      Grid World Value Matrix
*********************************************************************
  |   41.040094   |   47.991329   |   66.970499   |   78.766749   |
  -----------------------------------------------------------------
  |   36.035204   |   -1.263498   |   56.991329   |   66.970499   |
  -----------------------------------------------------------------
  |   36.594731   |   42.793026   |   48.736502   |   56.991329   |
  -----------------------------------------------------------------
  |   32.131959   |    0.000000   |   42.793026   |    0.000000   |
  -----------------------------------------------------------------
  |   28.213428   |   32.131959   |   36.594731   |   32.935258   |
  -----------------------------------------------------------------
```

Q Value

Observation

The Q Values may change during every subsequent execution of the program due to the randomization done at multiple levels.

- Choosing between Explore / Exploit
- Choosing a random action during Explore
- Environmental properties for couple of actions not being deterministic i.e., (right and up actions)

Nevertheless, the policy obtained remains consistent.

Other observations:

- Epsilon is initialized with 0.9 to favor more exploration during the early stages of learning.
- The Goal grid is made special, i.e., once inside the Goal grid, any action taken will lead to the Goal itself.

With Reward -1

```
*************************************************
            Grid World Reward Matrix
*************************************************
  |    -1    |    -1    |    -1    |    10    |
  ---------------------------------------------
  |    -1    |   -50    |    -1    |    -1    |
  ---------------------------------------------
  |    -1    |    -1    |    -1    |    -1    |
  ---------------------------------------------
  |    -1    |     0    |    -1    |     0    |
  ---------------------------------------------
  |    -1    |    -1    |    -1    |    -1    |
  ---------------------------------------------
```

```
     ********************************************************************************
                                Grid World Q Values Matrix
     ********************************************************************************
---------------------------------------------------------------------------------------------
|         61.97        |         69.97        |         78.86        |         99.73        |
|  61.97         69.98 |  61.97         78.87 |  69.97         88.75 |  99.73         99.73 |
|         54.77        |         17.87        |         69.97        |         99.73        |
---------------------------------------------------------------------------------------------
|         61.97        |         20.98        |         78.87        |         88.75        |
|  54.78         17.88 |   5.78         20.97 |  17.88         78.86 |  69.97         78.86 |
|         48.30        |          5.78        |         61.98        |         69.97        |
---------------------------------------------------------------------------------------------
|         54.78        |         17.88        |         69.98        |         78.86        |
|  48.30         54.78 |  48.30         61.98 |  54.78         69.97 |  61.98         69.97 |
|         42.47        |         54.78        |         54.78        |         69.97        |
---------------------------------------------------------------------------------------------
|         48.30        |          0.00        |         61.98        |          0.00        |
|  42.47         42.47 |   0.00          0.00 |  54.78         54.78 |   0.00          0.00 |
|         37.22        |          0.00        |         48.30        |          0.00        |
---------------------------------------------------------------------------------------------
|         42.47        |         42.47        |         54.78        |         42.47        |
|  37.22         42.47 |  37.22         48.30 |  42.47         42.47 |  48.30         42.47 |
|         37.22        |         42.47        |         48.30        |         42.47        |
---------------------------------------------------------------------------------------------
```

With Reward 0

```
*************************************************
            Grid World Reward Matrix
*************************************************
  |     0    |     0    |     0    |    10    |
  ---------------------------------------------
  |     0    |    -50   |     0    |     0    |
  ---------------------------------------------
  |     0    |     0    |     0    |     0    |
  ---------------------------------------------
  |     0    |     0    |     0    |     0    |
  ---------------------------------------------
  |     0    |     0    |     0    |     0    |
  ---------------------------------------------
```

```
     ********************************************************************************
                                Grid World Q Values Matrix
     ********************************************************************************
---------------------------------------------------------------------------------------------
|         61.33        |         68.11        |         75.77        |         94.09        |
|  61.33         68.18 |  61.30         75.83 |  68.10         84.60 |  94.28         94.08 |
|         55.18        |         16.40        |         68.12        |         94.30        |
---------------------------------------------------------------------------------------------
|         61.34        |         18.13        |         75.96        |         84.55        |
|  55.18         16.42 |   5.17         18.29 |  16.38         75.95 |  68.19         75.78 |
|         49.72        |          5.25        |         61.35        |         68.14        |
---------------------------------------------------------------------------------------------
|         55.17        |         16.42        |         68.30        |         75.92        |
|  49.73         55.27 |  49.72         61.44 |  55.25         68.25 |  61.37         68.24 |
|         44.75        |         55.26        |         55.25        |         68.25        |
---------------------------------------------------------------------------------------------
|         49.73        |          0.00        |         61.42        |          0.00        |
|  44.75         44.75 |   0.00          0.00 |  55.25         55.25 |   0.00          0.00 |
|         40.27        |          0.00        |         49.72        |          0.00        |
---------------------------------------------------------------------------------------------
|         44.75        |         44.75        |         55.26        |         44.71        |
|  40.28         44.75 |  40.27         49.73 |  44.75         44.72 |  49.71         44.72 |
|         40.27        |         44.75        |         49.72        |         44.72        |
---------------------------------------------------------------------------------------------
```