SOLVING MARKOV DECISION PROBLEMS

Artificial Intelligence and Machine learning

Nikodem Iwin 252928
Embedded Robotics
21.05.2023

Contents

1 About program

Program "mdp_soler.py" was written in Python 3.11.3, with help of libraries "numpy" and "matplotlib.pyplot". It solves the first part of the project "Part I: directly solving an MDP". To run the program you can just simply type: **python3 mdp_solver.py**

Be default it will run the "Russell and Norvig 4x3 world", which is saved in file "RN_world.txt" stored in folder "worlds". The programs read data and print the world based on the read values. Next it shows value of each state of the world, final values and optimal policy for moves. Program also displays the graph "Convergence of Utility Values", where on Y axis is utility and at X axis is Iterations.

The worlds used for testing program and writing this assignment are located in folder worlds and there should be:

- **assignment_world.txt**
- **assignment_world_b100.txt**
- **assignment_world_discount.txt**
- **assignment_world_p4.txt**
- **assignment_world_prob.txt**
- **RN_world.txt**

To change the worlds in program, I recommend simply uncommenting the specific name of file and commenting rest.

2 Russell and Norvig 4x3 world

This section presents the result of the default Russel and Norvig world, which is stored in "RN_world.txt" in folder "worlds". The console output was presented below. Please notice that in "Initialized Value Function:" forbidden state has value 0 and it isn't changed. Computed utilities were presented in "Final Value Function" and policy in "Optimal Policy". Utility values convergences after 16 iterations.

```
W World size:                          4 x 3
S Starting point:                      1 x 1
P Probabilities(p1, p2, p3, p4):       0.8, 0.1, 0.1, 0
R Reward:                              -0.04
G Discounting factor:                  1.0
T Terminal state:                      4, 2, -1.0
T Terminal state:                      4, 3, 1.0
F Prohibited states:                   2, 2
E Exploration parameter:               0
```

```
+---+---+---+---+
|   |   |   | T |
+---+---+---+---+
|   | F |   | T |
+---+---+---+---+
| S |   |   |   |
+---+---+---+---+
```

```
Initialized Value Function:
[[ 0.  0.  0.  1.]
 [ 0.  0.  0. -1.]
 [ 0.  0.  0.  0.]]

Final Value Function:
[[ 0.8116  0.8678  0.9178  1.    ]
 [ 0.7616  0.      0.6603 -1.    ]
 [ 0.7053  0.6553  0.6114  0.3879]]

Optimal Policy:
[['>' '>' '>' 'T']
 ['^' 'F' '^' 'T']
 ['^' '<' '^' '<']]
```

The graph "Convergence of Utility Values" for this world was presented at figure 1.

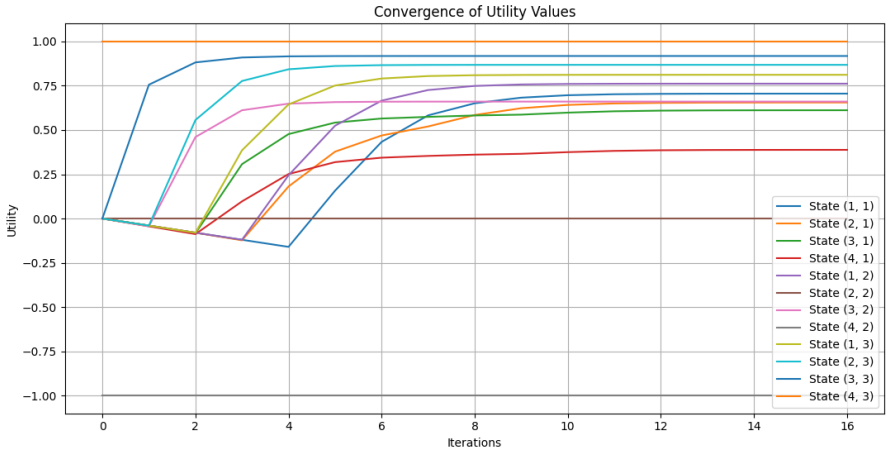

Figure 1: "Convergence of Utility Values" for Russell and Norvig 4x3 world

3 Assignment 4x4 world

This section presents the result of the assignment world, which is stored in "assignemnt_world.txt" in folder "worlds". The console output was presented below. Please notice that in "Initialized Value Function:" the forbidden state has value 0, the special state has value -20 and they aren't changed. Computed utilities were presented in "Final Value Function" and policy in "Optimal Policy". Utility values convergences after 28 iterations.

```
W World size:                          4 x 4
S Starting point:                      1 x 1
P Probabilities(p1, p2, p3, p4):       0.8, 0.1, 0.1, 0
R Reward:                              -1.0
G Discounting factor:                  0.99
T Terminal state:                      4, 1, 100.0
F Prohibited states:                   3, 1
B Special states:                      3, 2, -20.0
E Exploration parameter:               0.25
```

```
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
|   |   | B |   |
+---+---+---+---+
| S |   | F | T |
+---+---+---+---+
```

```
Initialized Value Function:
[[ 0.  0.  0.  0.]
```

```
[   0.    0.    0.    0.]
[   0.    0.  -20.    0.]
[   0.    0.    0.  100.]]

Final Value Function:
[[ 71.6296   74.0186   76.46     78.901  ]
 [ 69.8537   72.0653   74.7558   81.465  ]
 [ 67.5416   65.9687  -20.       84.5949]
 [ 65.2987   64.053     0.      100.     ]]

Optimal Policy:
[['>' '>' '>' 'v']
 ['>' '>' '>' 'v']
 ['^' '^' 'B' 'v']
 ['^' '^' 'F' 'T']]
```

The graph "Convergence of Utility Values" for this world was presented at figure 2.

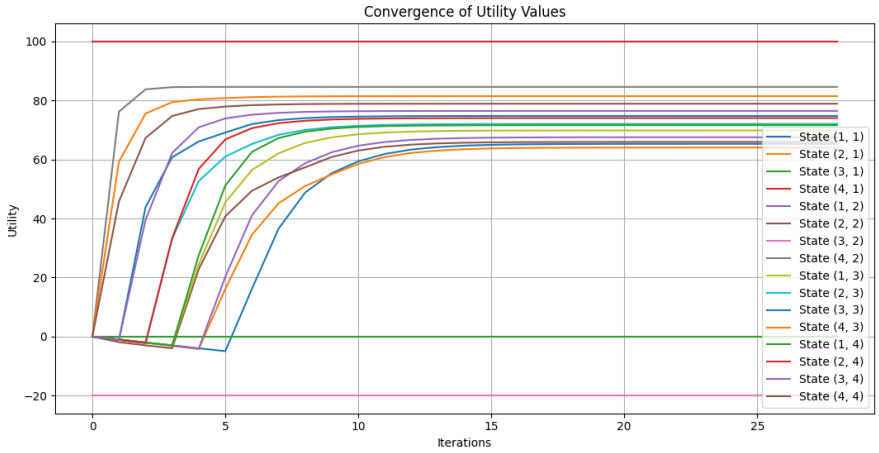

Figure 2: "Convergence of Utility Values" for assignment 4x4 world

4 Assignment 4x4 world with changed value of special state (100)

This section presents the result of the assignment world with changed value of special state from -20 to 100, which is stored in "assignemnt_world_b100.txt" in folder "worlds". The console output was presented below. Please notice that in "Initialized Value Function:" the forbidden state has value 0, the special state has value 100 and they aren't changed. Computed utilities were presented in "Final Value Function" and policy in "Optimal Policy". Utility values convergences after 16 iterations.

```
W World size:                          4 x 4
S Starting point:                      1 x 1
P Probabilities(p1, p2, p3, p4):       0.8, 0.1, 0.1, 0
R Reward:                              -1.0
G Discounting factor:                  0.99
T Terminal state:                      4, 1, 100.0
F Prohibited states:                   3, 1
B Special states:                      3, 2, 100.0
E Exploration parameter:               0.25
```

```
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
|   |   | B |   |
+---+---+---+---+
| S |   | F | T |
+---+---+---+---+
```

```
Initialized Value Function:
[[  0.    0.    0.    0.]
 [  0.    0.    0.    0.]
 [  0.    0.  100.    0.]
 [  0.    0.    0.  100.]]

Final Value Function:
[[ 89.9773  92.112   94.1801  93.186 ]
 [ 92.0895  94.546   97.0147  95.501 ]
 [ 93.9532  96.8817 100.      97.7802]
 [ 91.9806  94.158    0.     100.    ]]

Optimal Policy:
[['>' 'v' 'v' 'v']
 ['>' '>' 'v' 'v']
 ['>' '>' 'B' 'v']
```

['>' '^' 'F' 'T']]

The graph "Convergence of Utility Values" for this world was presented at figure 3.

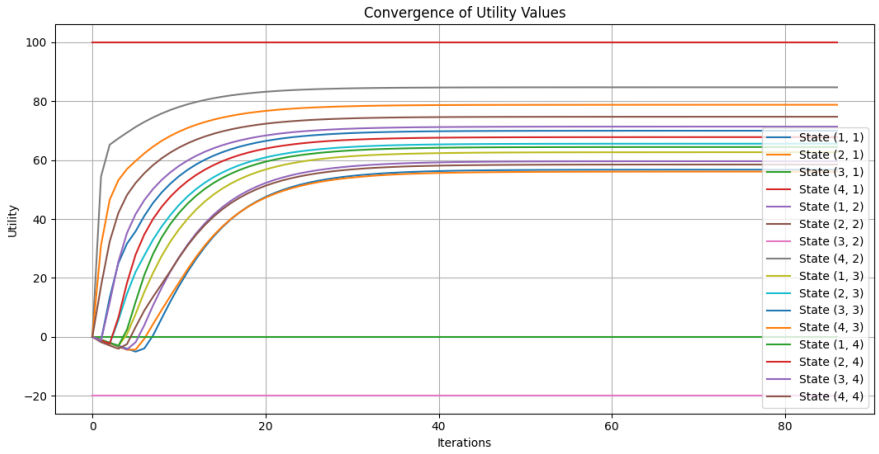

Figure 3: "Convergence of Utility Values" for assignment 4x4 world with changed special state value (100)

Because now the special state has equal value to the terminate state the policy has changed so the agent consider going closer to special state, what can be seen in states (1,1), (2,4), (3,3) and (3,4).

5 Assignment 4x4 world with changed probabilities

This section presents the result of the assignment world with changed probabilities values to p1 = 0.6, p2 = p3 = 0.2, which is stored in "assignemnt_world_prob.txt" in folder "worlds". The console output was presented below. Please notice that in "Initialized Value Function:" the forbidden state has value 0, the special state has value -20 and they aren't changed. Computed utilities were presented in "Final Value Function" and policy in "Optimal Policy". Utility values convergences after 86 iterations.

```
W World size:                          4 x 4
S Starting point:                      1 x 1
P Probabilities(p1, p2, p3, p4):       0.6, 0.2, 0.2, 0
R Reward:                              -1.0
G Discounting factor:                  0.99
T Terminal state:                      4, 1, 100.0
F Prohibited states:                   3, 1
B Special states:                      3, 2, -20.0
E Exploration parameter:               0.25
```

```
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
```

```
|   |   |   |   |
+---+---+---+---+
|   |   | B |   |
+---+---+---+---+
| S |   | F | T |
+---+---+---+---+
```

```
Initialized Value Function:
[[  0.    0.    0.    0.]
 [  0.    0.    0.    0.]
 [  0.    0.  -20.    0.]
 [  0.    0.    0.  100.]]

Final Value Function:
[[ 64.4468  67.8059  71.3739  74.7239]
 [ 62.6753  65.5769  69.9792  78.7821]
 [ 59.617   58.5037 -20.      84.7262]
 [ 56.7574  56.0962   0.     100.    ]]

Optimal Policy:
[['>' '>' '>' 'v']
 ['>' '>' '>' 'v']
 ['^' '^' 'B' 'v']
 ['^' '^' 'F' 'T']]
```

The graph "Convergence of Utility Values" for this world was presented at figure 4.

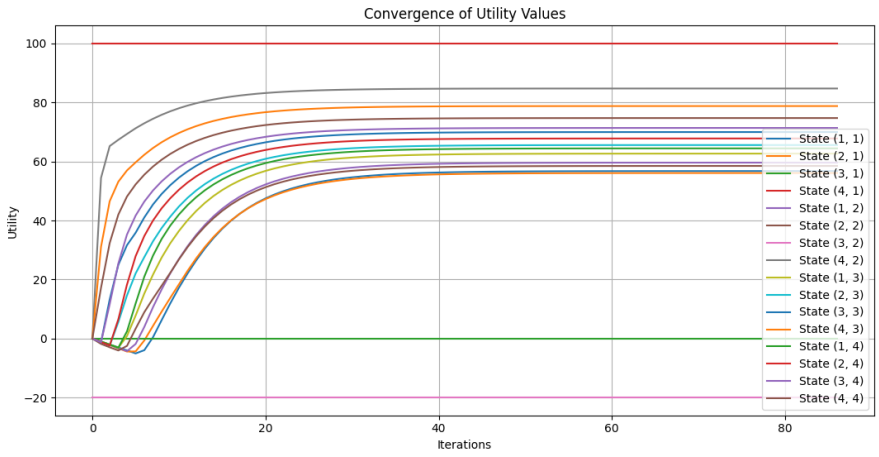

Figure 4: "Convergence of Utility Values" for assignment 4x4 world with changed probabilities

Changing the the values in the way that p1 is still bigger, and p2 is equal to p3 changes only utility values, however the agent still makes the same moves.

6 Assignment 4x4 world with possibility to return (p4 != 0)

This section presents the result of the assignment world with changed probabilities so there is p4, which is stored in "assignemnt_world_p4.txt" in folder "worlds". The console output was presented below. Please notice that in "Initialized Value Function:" the forbidden state has value 0, the special state has value -20 and they aren't changed. Computed utilities were presented in "Final Value Function" and policy in "Optimal Policy". Utility values convergences after 54 iterations.

```
W World size:                          4 x 4
S Starting point:                      1 x 1
P Probabilities(p1, p2, p3, p4):       0.6, 0.1, 0.1, 0.20000000000000004
R Reward:                              -1.0
G Discounting factor:                  0.99
T Terminal state:                      4, 1, 100.0
F Prohibited states:                   3, 1
B Special states:                      3, 2, -20.0
E Exploration parameter:               0.25
```

```
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
|   |   | B |   |
+---+---+---+---+
| S |   | F | T |
+---+---+---+---+
```

```
Initialized Value Function:
[[   0.    0.    0.    0.]
 [   0.    0.    0.    0.]
 [   0.    0.  -20.    0.]
 [   0.    0.    0.  100.]]
```

```
Final Value Function:
[[ 54.6253  57.8579  62.5444  67.9184]
 [ 50.8483  52.1497  56.0923  71.6409]
 [ 46.4852  40.7577 -20.      78.3628]
 [ 43.6579  41.206    0.     100.    ]]
```

```
Optimal Policy:
[['>' '>' '>' 'v']
 ['^' '^' '>' 'v']
```

```
['^'  '^'  'B'  'v']
['^'  '<'  'F'  'T']]
```

The graph "Convergence of Utility Values" for this world was presented at figure 5.

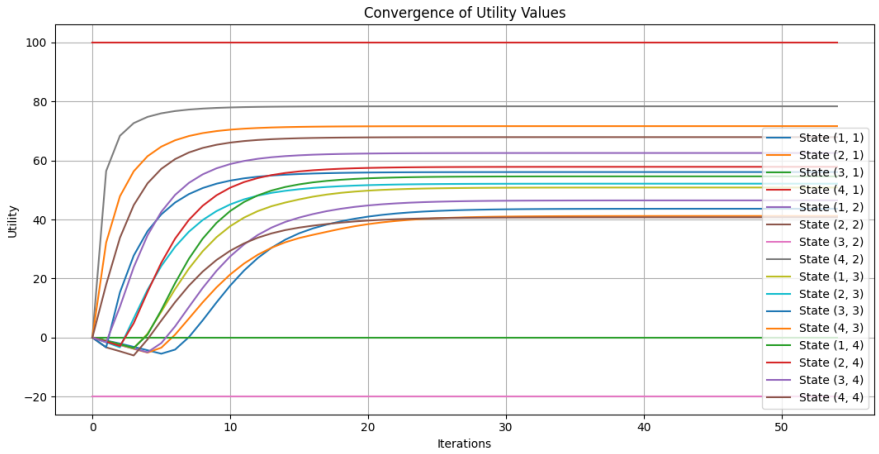

Figure 5: "Convergence of Utility Values" for assignment 4x4 world with possibility to return (p4 != 0)

Because now there is probability p4 equal to 0.2 the agent is more likely to return than in default assignment world, so it try to stay away of forbidden state what is represented in state (2,1).

7 Assignment 4x4 world with changed discounted factor (0.5)

This section presents the result of the assignment world with changed discount factor value, which is stored in "assignemnt_world_discount.txt" in folder "worlds". The console output was presented below. Please notice that in "Initialized Value Function:" the forbidden state has value 0, the special state has value -20 and they aren't changed. Computed utilities were presented in "Final Value Function" and policy in "Optimal Policy". Utility values convergences after 11 iterations.

```
W World size:                          4 x 4
S Starting point:                      1 x 1
P Probabilities(p1, p2, p3, p4):       0.8, 0.1, 0.1, 0
R Reward:                              -1.0
G Discounting factor:                  0.5
T Terminal state:                      4, 1, 100.0
F Prohibited states:                   3, 1
B Special states:                      3, 2, -20.0
E Exploration parameter:               0.25
```

```
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
|   |   |   |   |
+---+---+---+---+
|   |   | B |   |
+---+---+---+---+
| S |   | F | T |
+---+---+---+---+
```

```
Initialized Value Function:
[[  0.    0.    0.    0.]
 [  0.    0.    0.    0.]
 [  0.    0.  -20.    0.]
 [  0.    0.    0.  100.]]

Final Value Function:
[[ -1.2388  -0.3349   1.6175   5.7802]
 [ -0.8583   0.6968   4.4912  16.0259]
 [ -1.5013  -1.6579 -20.     40.    ]
 [ -1.7818  -1.8445   0.    100.    ]]

Optimal Policy:
[['>' '>' '>' 'v']
 ['>' '>' '>' 'v']
 ['^' '^' 'B' 'v']
 ['^' '>' 'F' 'T']]
```

The graph "Convergence of Utility Values" for this world was presented at figure 6.

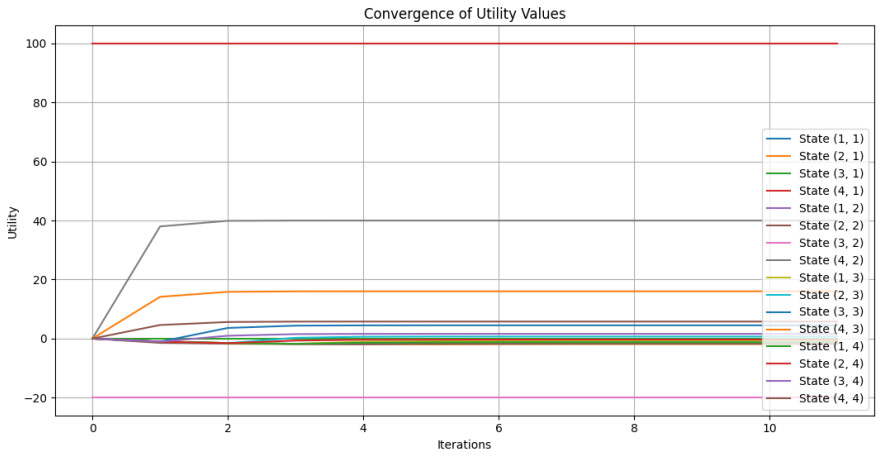

Figure 6: "Convergence of Utility Values" for assignment 4x4 world with changed discounted factor (0.5)

Lowering the discounting factor changed the values of the state so, they aren't as important as before. That result in very fast computation and stopping after 11 iteration, when with default factor (0.99) it was 27.

8 Hyper world

This section presents the result of the hyper world, which is stored in "hyper_world.txt" in folder "worlds". The console output was presented below. Please notice that in "Initialized Value Function:" the terminal states have value 25, -25, the special states have value -25, 25, the forbidden states have value 0 and they aren't changed. Computed utilities were presented in "Final Value Function" and policy in "Optimal Policy". Utility values convergences after 35 iterations.

```
W World size:                          5 x 5
S Starting point:                      3 x 3
P Probabilities(p1, p2, p3, p4):       0.5, 0.25, 0.25, 0
R Reward:                              -1.0
G Discounting factor:                  0.99
T Terminal state:                      1, 5, -25.0
T Terminal state:                      1, 1, 25.0
F Prohibited states:                   2, 2
F Prohibited states:                   2, 4
F Prohibited states:                   4, 2
F Prohibited states:                   4, 4
B Special states:                      5, 5, 25.0
B Special states:                      5, 1, -25.0
E Exploration parameter:               0.25
```

```
+---+---+---+---+---+
| T |   |   |   | B |
+---+---+---+---+---+
|   | F |   | F |   |
+---+---+---+---+---+
|   |   | S |   |   |
+---+---+---+---+---+
|   | F |   | F |   |
+---+---+---+---+---+
| T |   |   |   | B |
+---+---+---+---+---+
```

```
Initialized Value Function:
[[-25.   0.   0.   0.  25.]
 [  0.   0.   0.   0.   0.]
 [  0.   0.   0.   0.   0.]
 [  0.   0.   0.   0.   0.]
 [ 25.   0.   0.   0. -25.]]
```

```
Final Value Function:
[[-25.      16.5887  18.9441  22.5248  25.    ]
 [ 16.5888   0.      16.5887   0.      22.5248]
 [ 18.9441  16.5888  15.4228  16.5887  18.9441]
 [ 22.5248   0.      16.5888   0.      16.5887]
 [ 25.      22.5248  18.9441  16.5888 -25.    ]]

Optimal Policy:
[['T' '>' '>' '>' 'B']
 ['v' 'F' '^' 'F' '^']
 ['v' '<' 'v' '>' '^']
 ['v' 'F' 'v' 'F' '^']
 ['T' '<' '<' '<' 'B']]
```

The graph "Convergence of Utility Values" for this world was presented at figure 7.

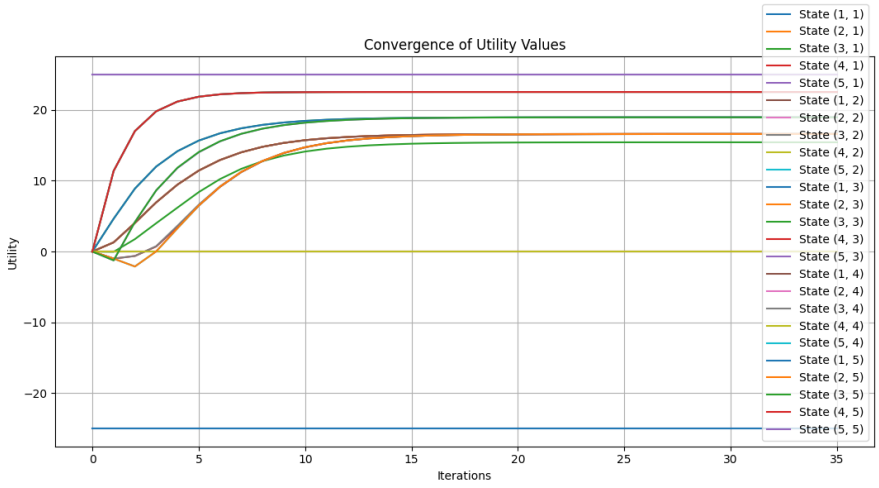

Figure 7: "Convergence of Utility Values" for assignment 4x4 world with changed discounted factor (0.5)

This world is more complex than the previous one, because of multiple states, with positive reward. It was made with fought of "symmetry" which is shown by the policy, what confirms of correct operation of the program.

9 Labyrinth

This section presents the result of the "labyrinth" world, which is stored in "labyrinth.txt" in folder "worlds". The console output was presented below. Please notice that in "Initialized Value Function:" the special state have value 25, the forbidden states have value 0 and they aren't changed. Computed utilities were presented in "Final Value Function" and policy in "Optimal Policy". Utility values convergences after 979 iterations.

```
W World size:                          10 x 10
S Starting point:                      1 x 1
P Probabilities(p1, p2, p3, p4):       0.25, 0.25, 0.25, 0.25
R Reward:                              -0.5
G Discounting factor:                  1.0
T Terminal state:                      10, 8, 25.0
F Prohibited states:                   1, 4
F Prohibited states:                   1, 8
F Prohibited states:                   1, 9
F Prohibited states:                   1, 10
F Prohibited states:                   2, 1
F Prohibited states:                   2, 2
F Prohibited states:                   2, 4
F Prohibited states:                   2, 6
F Prohibited states:                   2, 10
F Prohibited states:                   3, 6
F Prohibited states:                   3, 7
F Prohibited states:                   3, 8
F Prohibited states:                   4, 2
F Prohibited states:                   4, 6
F Prohibited states:                   4, 9
F Prohibited states:                   5, 2
F Prohibited states:                   5, 8
F Prohibited states:                   6, 2
F Prohibited states:                   6, 3
F Prohibited states:                   6, 4
F Prohibited states:                   6, 5
F Prohibited states:                   6, 6
F Prohibited states:                   7, 2
F Prohibited states:                   7, 3
F Prohibited states:                   8, 5
F Prohibited states:                   8, 9
F Prohibited states:                   9, 1
F Prohibited states:                   9, 3
F Prohibited states:                   9, 4
F Prohibited states:                   9, 5
F Prohibited states:                   9, 6
F Prohibited states:                   9, 7
```

```
F Prohibited states:                    9, 8
F Prohibited states:                    10, 7
F Prohibited states:                    10, 10
E Exploration parameter:                0.25
```

```
+---+---+---+---+---+---+---+---+---+---+---+
| F | F |   |   |   |   |   |   |   |   | F |
+---+---+---+---+---+---+---+---+---+---+---+
| F |   |   |   | F |   |   |   | F |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
| F |   |   | F |   |   | F |   |   |   | F | T |
+---+---+---+---+---+---+---+---+---+---+---+
|   |   |   | F |   |   |   |   |   |   | F | F |
+---+---+---+---+---+---+---+---+---+---+---+
|   |   | F | F | F |   |   | F |   |   | F |   |
+---+---+---+---+---+---+---+---+---+---+---+
|   |   |   |   |   |   |   | F |   | F | F |   |
+---+---+---+---+---+---+---+---+---+---+---+
| F | F |   |   |   |   |   | F |   |   | F |   |
+---+---+---+---+---+---+---+---+---+---+---+
|   |   |   |   |   |   |   | F | F |   | F |   |
+---+---+---+---+---+---+---+---+---+---+---+
|   |   | F |   |   | F | F | F | F |   |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
| S | F |   |   |   |   |   |   |   |   | F |   |
+---+---+---+---+---+---+---+---+---+---+---+
```

```
Initialized Value Function:
[[ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0. 25.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]]
```

```
Final Value Function:
[[   0.         0.       −25.2181   −27.7251   −28.2321   −27.5124   −25.4283
   −21.504    −15.5798     0.     ]
 [   0.       −15.6973   −20.7112     0.       −27.4588   −26.8768   −25.2685
     0.        −7.6556     7.6722]
 [   0.       −24.3807     0.       −19.2181     0.       −25.2676   −21.5003
```

```
   −13.0592     0.        25.      ]
[ −65.7996   −55.4449     0.      −27.0513   −28.6032   −25.4256   −20.4061
   −15.6772     0.        0.      ]
[ −74.1544     0.        0.        0.       −31.3326     0.       −17.0214
   −11.5662     0.      −15.0776]
[ −80.5094   −84.8647   −84.199   −77.8464   −63.3947     0.       −17.0919
     0.         0.      −28.1552]
[   0.         0.      −87.8859   −83.9455   −79.0052     0.       −32.2544
   −27.9814     0.      −39.2328]
[−104.8776   −98.8785   −93.5134   −89.0448   −87.6755     0.         0.
   −49.6899     0.      −48.3105]
[−108.8769     0.      −96.2447     0.         0.         0.         0.
   −60.6222   −58.4662   −55.3883]
[−110.8765     0.      −97.1551   −96.0658   −92.9768   −87.8879   −80.7992
   −71.7106     0.      −57.3882]]
```

Optimal Policy:

```
[['F' 'F' '<' 'v' 'v' '>' '>' 'v' '>' 'F']
 ['F' '^' 'v' 'F' 'v' 'v' '>' 'F' '>' 'v']
 ['F' '<' 'F' '^' 'F' '<' '>' '^' 'F' 'T']
 ['^' 'v' 'F' 'v' '^' 'v' '>' '>' 'F' 'F']
 ['>' 'F' 'F' 'F' '<' 'F' '<' 'v' 'F' '^']
 ['v' '^' '^' '^' '>' 'F' '<' 'F' 'F' '<']
 ['F' 'F' '<' '^' '>' 'F' 'v' '^' 'F' '<']
 ['^' '^' '^' 'v' 'v' 'F' 'F' '<' 'F' '<']
 ['>' 'F' '<' 'F' 'F' 'F' 'F' '<' '^' '^']
 ['>' 'F' '<' '^' '^' '^' '^' '>' 'F' '<']]
```

The graph "Convergence of Utility Values" for this world was presented at figure 8.

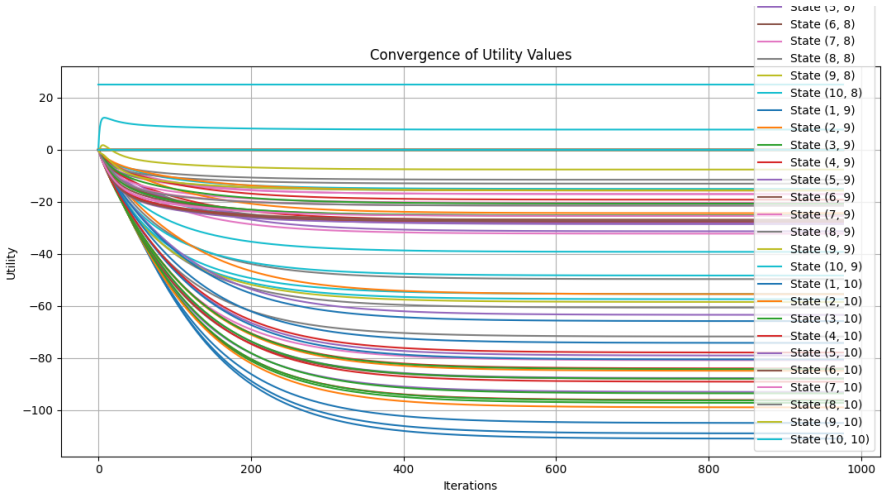

Figure 8: "Convergence of Utility Values" for assignment 4x4 world with changed discounted factor (0.5)

As you can see the program, can solve the problem, however the solution makes agent to face the wall many times and it is the result of equal probability of each move (Up, Down, Left and Right). If you change the probability to 0.8, 0.1 and 0.1 it makes agent to be more concern about main move, than the values connects to p2 and p3. These situation was presented below by world "labyrinth2.txt", stored in "worlds" and by figure 9. It also solved the map in only 64 iterations.

```
W World size:                          10 x 10
S Starting point:                      1 x 1
P Probabilities(p1, p2, p3, p4):       0.8, 0.1, 0.1, 0
R Reward:                              -0.5
G Discounting factor:                  1.0
T Terminal state:                      10, 8, 25.0
F Prohibited states:                   1, 4
F Prohibited states:                   1, 8
F Prohibited states:                   1, 9
F Prohibited states:                   1, 10
F Prohibited states:                   2, 1
F Prohibited states:                   2, 2
F Prohibited states:                   2, 4
F Prohibited states:                   2, 6
F Prohibited states:                   2, 10
F Prohibited states:                   3, 6
F Prohibited states:                   3, 7
F Prohibited states:                   3, 8
F Prohibited states:                   4, 2
F Prohibited states:                   4, 6
F Prohibited states:                   4, 9
F Prohibited states:                   5, 2
F Prohibited states:                   5, 8
F Prohibited states:                   6, 2
F Prohibited states:                   6, 3
F Prohibited states:                   6, 4
F Prohibited states:                   6, 5
F Prohibited states:                   6, 6
F Prohibited states:                   7, 2
F Prohibited states:                   7, 3
F Prohibited states:                   8, 5
F Prohibited states:                   8, 9
F Prohibited states:                   9, 1
F Prohibited states:                   9, 3
F Prohibited states:                   9, 4
F Prohibited states:                   9, 5
F Prohibited states:                   9, 6
F Prohibited states:                   9, 7
F Prohibited states:                   9, 8
```

```
F Prohibited states:                    10, 7
F Prohibited states:                    10, 10
E Exploration parameter:                0.25
```

```
+---+---+---+---+---+---+---+---+---+---+---+
| F | F |   |   |   |   |   |   |   |   | F |
+---+---+---+---+---+---+---+---+---+---+---+
| F |   |   | F |   |   |   | F |   |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
| F |   | F |   | F |   |   |   | F | T |   |
+---+---+---+---+---+---+---+---+---+---+---+
|   |   | F |   |   |   |   |   | F | F |   |
+---+---+---+---+---+---+---+---+---+---+---+
|   | F | F | F |   | F |   |   | F |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
|   |   |   |   |   | F |   | F | F |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
| F | F |   |   |   | F |   |   | F |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
|   |   |   |   |   | F | F |   | F |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
|   | F |   | F | F | F | F |   |   |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
| S | F |   |   |   |   |   |   | F |   |   |
+---+---+---+---+---+---+---+---+---+---+---+
```

```
Initialized Value Function:
[[ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0. 25.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]]

Final Value Function:
[[ 0.       0.      17.9956 18.7407 19.3657 20.0528 20.7551 21.4679 22.0929
   0.     ]
 [ 0.      14.3475 17.0347  0.     18.8689 19.4344 20.0528  0.     22.7961
  24.1996]
 [ 0.      12.1978  0.      15.8674  0.     18.8477 19.1542 17.2711
0.
```

```
  25.     ]
 [10.8468 11.4921  0.      16.3251 17.2057 18.1391 18.4033 17.6608
0.
   0.     ]
 [10.6844  0.      0.       0.      14.7384  0.      17.7041 17.11
0.
   6.4218]
 [11.3094 12.0125 12.629  13.3225 14.0255  0.      15.1814  0.
0.
   7.0468]
 [ 0.      0.      12.0811 12.6983 13.3225  0.      14.2773 12.0447
0.
   7.6718]
 [10.1074 10.8105 11.454  12.0811 12.629   0.       0.      11.229
0.
   8.2968]
 [ 9.4823  0.      10.663   0.       0.       0.       0.      10.5039
9.7031
   8.9218]
 [ 8.8572  0.       9.9598  9.3348  8.7097  8.5507  9.1757  9.8007
0.
   8.2968]]

Optimal Policy:
[['F' 'F' '>' '>' '>' '>' '>' '>' 'v' 'F']
 ['F' '>' '^' 'F' '>' '^' '^' 'F' '>' 'v']
 ['F' '^' 'F' 'v' 'F' '^' '^' '<' 'F' 'T']
 ['>' '^' 'F' '>' '>' '^' '^' '<' 'F' 'F']
 ['v' 'F' 'F' 'F' '^' 'F' '^' '<' 'F' 'v']
 ['>' '>' '>' '>' '^' 'F' '^' 'F' 'F' 'v']
 ['F' 'F' '>' '^' '^' 'F' '^' '<' 'F' 'v']
 ['>' '>' '^' '^' '^' 'F' 'F' '^' 'F' 'v']
 ['^' 'F' '^' 'F' 'F' 'F' 'F' '^' '<' '<']
 ['^' 'F' '^' '<' '<' '>' '>' '^' 'F' '^']]
```

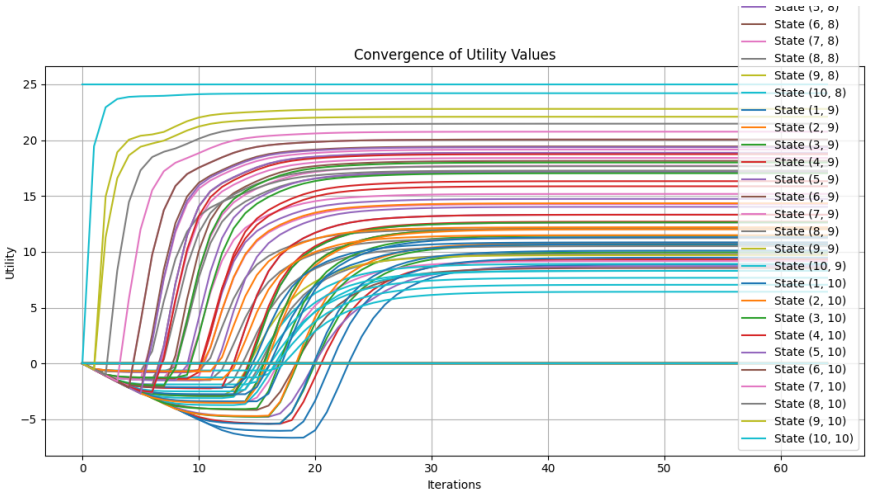

Figure 9: "Convergence of Utility Values" for assignment 4x4 world with changed discounted factor (0.5)