Markov Decision Problem

Jacek Multan, index: 248964

June 2022

1 Directly Solving MDP

1.1 Solution of the 4x3 problem

The problem is given by parameters;

- World size 4x3
- Probabilities p1 = 0,8; p2 = 0,1; p3 = 0,1
- Default reward = -0,04
- Discounting factor γ = 1,00

The World with calculated utilities and policy is shown on the figure 1. Symbol "F" means the tile is forbidden (for example there might be some obstacle). Symbol "T" means terminal state. First line contains utility of the tile, second is reward and third is optimal move. The results are the same as given during lecture, so the algorithm is most likely working correctly. The algorithm run 24 iterations.

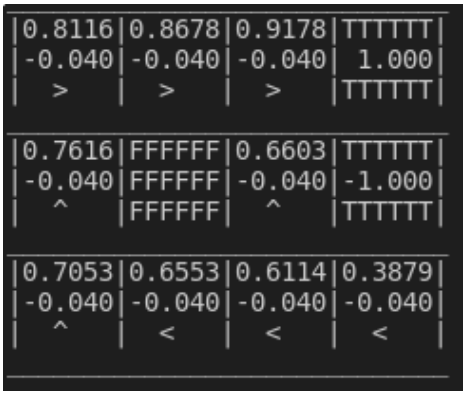

Figure 1: Calculated World 1

Convergance plot is presented on figure 2.

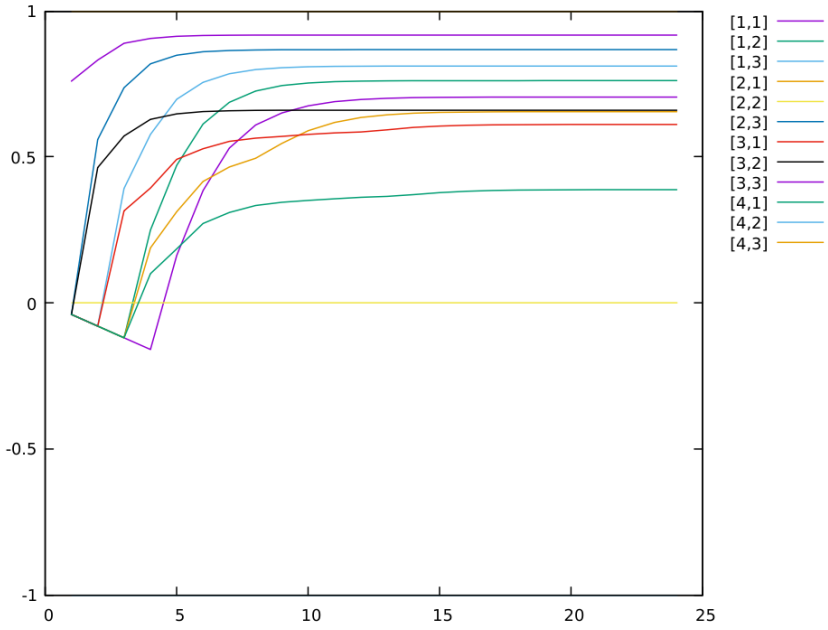

Figure 2: Convergance graph for world 1

1.2 World 4x4

The problem is given by parameters;

- World size 4x4
- Probabilities p1 = 0,8; p2 = 0,1; p3 = 0,1
- Default reward = -1
- Discounting factor $\gamma = 0{,}99$

The World with calculated utilities and policy is shown on the figure 3. Symbols and numbers have the same meaning as in the previous section. The algorithm was run with 36 iterations.

```
|81.9384|84.2610|86.5861|88.8827|
|-1.000|-1.000|-1.000|-1.000|
|  >   |  >   |  >   |  v   |
_________________________________
|81.7354|84.2724|87.0596|91.5547|
|-1.000|-1.000|-1.000|-1.000|
|  >   |  >   |  >   |  v   |
_________________________________
|79.5936|80.5997|70.4670|94.5352|
|-1.000|-1.000|-20.000|-1.000|
|  ^   |  ^   |  >   |  v   |
_________________________________
|77.4526|78.2495|FFFFFF|TTTTTT|
|-1.000|-1.000|FFFFFF| 100.000|
|  ^   |  ^   |FFFFFF|TTTTTT|
_________________________________
```

Figure 3: Calculated World 2

Convergance plot is presented on figure 4.

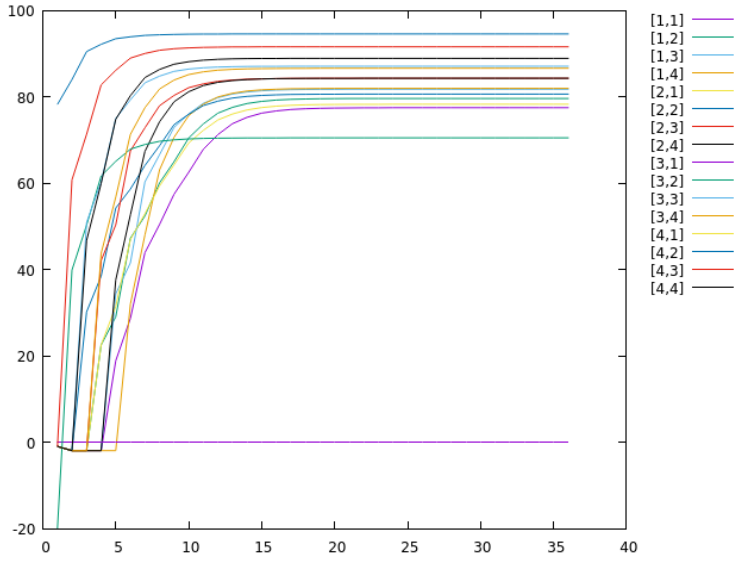

Figure 4: Convergance graph for world 2

1.3 World 4x4 with modified rewards for termianl state, each move and special state

During this task rewards were changed:

- Default reward = -2
- Terminal state = 90
- Special state = -90

The World with calculated utilities and policy is shown on the figure 5. Symbols and numbers have the same meaning as in the previous section.

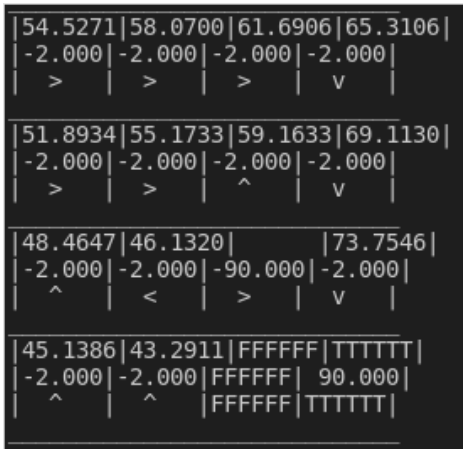

Figure 5: Modified rewards for World 2

Convergance plot is presented on figure 6.

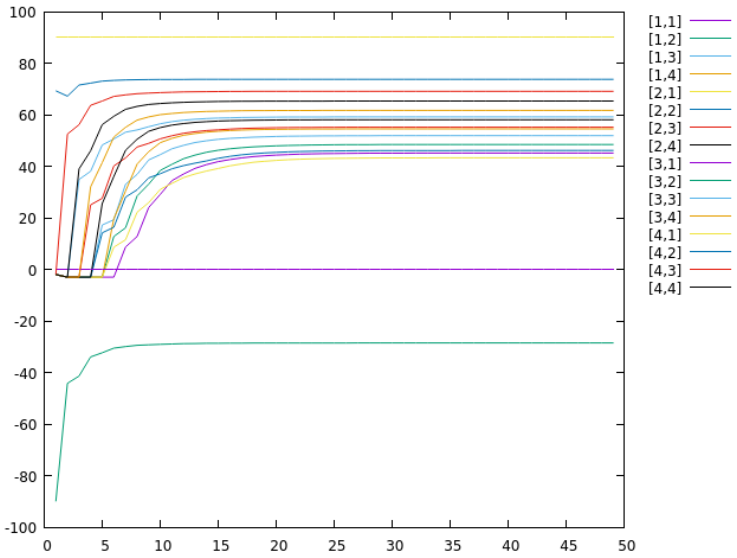

Figure 6: Convergance graph for world 2 with modified rewards

1.3.1 The results

The penalty for entering the special tile was significantly higher this time, so optimal actions were doing everything possible to avoid it by moving the other direction. The only exception is close to the terminal state as move cost was increased too. It's better to risk moving into the special state than tring to bump into the right direction moving againt the wall. The utility of special state is negative because of the very large penalty. It can be treated as a very bad state.

1.4 World 4x4 with modified uncertainity model

During this task parameters of the second world were modified:

- Probabilities p1 = 0,6; p2 = 0,1; p3 = 0,3

The World with calculated utilities and policy is shown on the figure 7. Symbols and numbers have the same meaning as in the previous section.

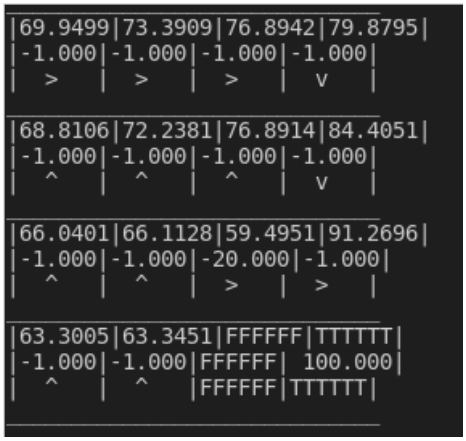

Figure 7: Modified action uncertainity for World 2

Convergance plot is presented on figure 8.

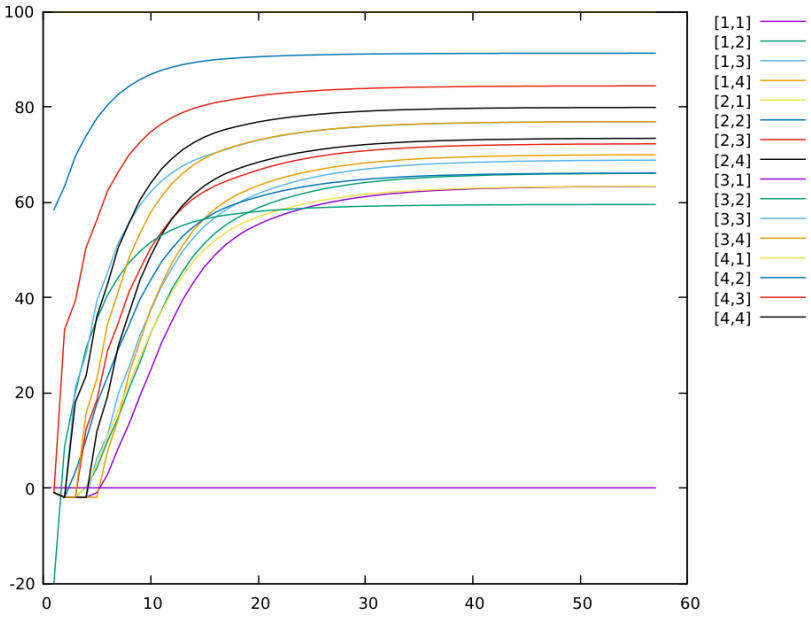

Figure 8: Convergance graph for world 2 with modified action uncertainity

1.4.1 The results

By increasing the probability to move right from the desired state, the agent shouldn't perfom move in which she has special state on her right unless it would produce a lot of extra moves (like in the tile (2,2)). Close to the termianl state it's more sensible to bump against wall than go straight for the reward.

1.5 World 4x4 with modified discounting factor

During this task parameters of the second world were modified:

- Discounting factor $\gamma = 0{,}92$

The World with calculated utilities and policy is shown on the figure 9. Symbols and numbers have the same meaning as in the previous section.

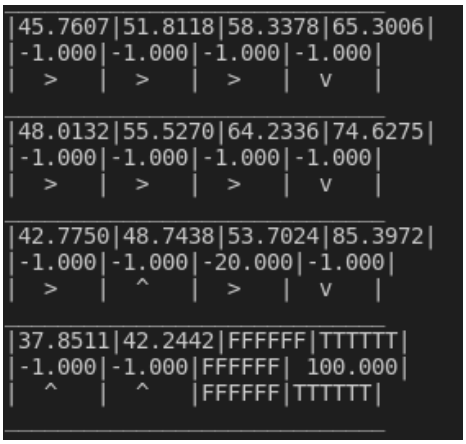

Figure 9: Modified discounting factor for World 2

Convergance plot is presented on figure 10.

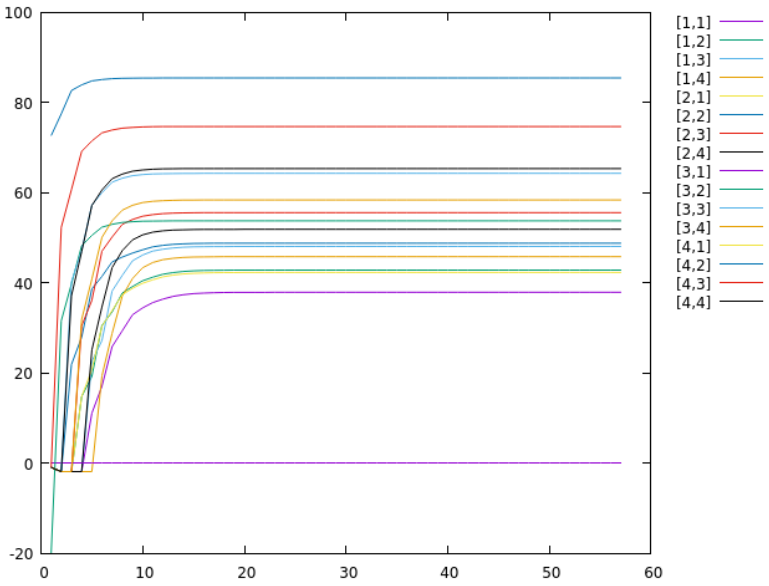

Figure 10: Convergance graph for world 2 with modified discounting factor

1.5.1 The results

In this case changing the discounting factor didn't change moving policy a lot, but the utilities have changed significantly. There is visible slight tendency to finish the world as quick as possible, as previous moves are forgotten quickly.