AI Assignment 2 PART B - Team 4

Prathyakshun Rajashankar 20161107

Anish Gulati 20161213

Shown below are the policies and the utilities of each of the 16 states in each iteration.

The value iteration algorithm when run for X = 4 (Team number) ran 15 iterations after which the utilities for all the states changed less than 1%.

(1, 2) cell contains the **WALL**, hence the utilities are not meantioned there.

(0, 0) and (3, 3) contains the terminal states.

Policy for iter ation 1

	0	**1**	**2**	**3**
0	0.0	R	R	R
1	D	R	**W**	R
2	U	U	R	D
3	L	L	R	0.0

Utility for iter ation 1

	0	**1**	**2**	**3**
0	-4.0	-0.8	-0.8	-0.8
1	-0.48	-0.8	**W**	-0.8
2	-0.08	-0.52	-0.8	2.4
3	-0.52	-1.2	2.4	4.0

In the top row, the policy says to go rightwards as it moves away from the most negative state -4 (0, 0).

(2, 0) shows up for now as it is just the 1st iteration and is expected to change.

Policy for iter ation 2

	0	1	2	3
0	0.0	R	R	R
1	D	U	**W**	D
2	U	U	D	D
3	L	R	R	0.0

Utility for iteration 2

	0	1	2	3
0	-4.0	-1.6	-1.6	-1.6
1	-0.992	-1.316	**W**	0.96
2	-0.564	-1.064	1.308	2.56
3	-1.036	0.548	2.56	4.0

(1, 3) state now tends to go down as the negative value at (0,3) became more negative and (2, 3) has become more positive. Since negativity increased in (1,1) the state at (1, 1) prefers going up. However, the direction is expected to change as it is just the 2nd iteration.

Policy for iteration 3

	0	1	2	3
0	0.0	R	R	D
1	D	U	**W**	D
2	U	R	R	D
3	R	R	R	0.0

Utility for iteration 3

	0	**1**	**2**	**3**
0	-4.0	-2.372	-2.4	-0.352
1	-1.482	-1.86	**W**	1.44
2	-1.054	0.17	1.635	2.787
3	-0.522	0.796	2.787	4.0

(3, 3) is expected to go down as the (2, 3) has become more positive. Similarly (3, 0) direction is changed to right as the positiveness in (3, 1) has increased. Same is the case with (2, 1) as the positiveness at (2, 2) has increased.

States 2 levels away from the negative sink are getting more and more negative due to the influence of the negative sink and the increase in negativity of the neighbors. This is expected to stop in the later stage once the other neighbors start getting positive.

Policy for iteration 4

	0	**1**	**2**	**3**
0	0.0	D	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 4

	0	**1**	**2**	**3**
0	-4.0	-2.928	-1.562	0.077
1	-1.977	-0.999	**W**	1.717
2	-0.465	0.401	1.872	2.842
3	-0.32	1.126	2.842	4.0

(2, 0) direction changed to the positive sink value because of the positive value at (2, 1)

(1, 1) has changed its direction downwards due to the occurrence of the positive value at (2, 1)

The values at (1, 0) and (1, 1) have become more negative due to the influence of the negative sink -4 at (0, 0)

Policy for iteration 5

	0	**1**	**2**	**3**
0	0.0	D	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 5

	0	**1**	**2**	**3**
0	-4.0	-2.155	-1.051	0.425
1	-1.469	-0.776	**W**	1.817
2	-0.309	0.71	1.945	2.871
3	0.022	1.226	2.871	4.0

The values at (1, 0) and (1, 1) have stopped becoming more negative due to the other neighbors apart from the negative sink (0, 0) which have gotten less negative.

(2, 0) is less negative than (1, 1) due to the presence of positive reward 0.4 at (2, 0) state.

Policy for iteration 6

	0	**1**	**2**	**3**
0	0.0	D	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 6

	0	**1**	**2**	**3**
0	-4.0	-1.926	-0.67	0.591
1	-1.271	-0.457	**W**	1.861
2	0.023	0.801	1.979	2.882
3	0.153	1.291	2.882	4.0

The directions remain the same. We see that the (0, 3) is more positive compared to the (3, 0) state because of the the influence of the negative reward -0.4 in (3, 1) state.

Policy for iteration 7

	0	**1**	**2**	**3**
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

(0, 1) has changed its direction to right as (0, 2) negativity reduced slightly more than (1, 1).

Now all the states have started pointing in the direction of the postive hole (3, 3). The state values keep getting more and more positive till the algorithm terminates. The algorithm here terminates when the change in the value is less than 1%(tolerance) in all the states in an iteration. The directions remain constant throughout from now on.

Utility for iteration 7

	0	**1**	**2**	**3**
0	-4.0	-1.574	-0.461	0.681
1	-0.954	-0.332	**W**	1.877
2	0.129	0.866	1.991	2.886
3	0.25	1.315	2.886	4.0

Policy for iteration 8

	0	**1**	**2**	**3**
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 8

	0	**1**	**2**	**3**
0	-4.0	-1.359	-0.348	0.724
1	-0.825	-0.235	**W**	1.884
2	0.223	0.891	1.997	2.888
3	0.29	1.327	2.888	4.0

Policy for iteration 9

	0	**1**	**2**	**3**
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 9

	0	**1**	**2**	**3**
0	-4.0	-1.238	-0.29	0.745
1	-0.728	-0.193	**W**	1.887
2	0.259	0.906	1.999	2.888
3	0.313	1.332	2.888	4.0

Policy for iteration 10

	0	**1**	**2**	**3**
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 10

	0	**1**	**2**	**3**
0	-4.0	-1.175	-0.262	0.755
1	-0.685	-0.167	**W**	1.888
2	0.284	0.913	1.999	2.889
3	0.323	1.335	2.889	4.0

Policy for iteration 11

	0	**1**	**2**	**3**
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 11

	0	**1**	**2**	**3**
0	-4.0	-1.144	-0.248	0.76
1	-0.658	-0.155	**W**	1.889
2	0.294	0.916	2.0	2.889
3	0.328	1.336	2.889	4.0

Policy for iteration 12

	0	**1**	**2**	**3**
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 12

	0	**1**	**2**	**3**
0	-4.0	-1.129	-0.242	0.762
1	-0.646	-0.148	**W**	1.889
2	0.3	0.918	2.0	2.889
3	0.331	1.336	2.889	4.0

Policy for iteration 13

	0	**1**	**2**	**3**
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 13

	0	1	2	3
0	-4.0	-1.121	-0.239	0.763
1	-0.639	-0.145	**W**	1.889
2	0.303	0.919	2.0	2.889
3	0.332	1.337	2.889	4.0

Policy for iteration 14

	0	1	2	3
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 14

	0	1	2	3
0	-4.0	-1.118	-0.237	0.764
1	-0.636	-0.143	**W**	1.889
2	0.304	0.919	2.0	2.889
3	0.333	1.337	2.889	4.0

Policy for iteration 15

	0	1	2	3
0	0.0	R	R	D
1	D	D	**W**	D
2	R	R	R	D
3	R	R	R	0.0

Utility for iteration 15

	0	**1**	**2**	**3**
0	-4.0	-1.116	-0.237	0.764
1	-0.635	-0.143	**W**	1.889
2	0.305	0.919	2.0	2.889
3	0.333	1.337	2.889	4.0

	0	**1**	**2**	**3**
0	-4.0	-1.116	-0.237	0.764
1	-0.635	-0.143	**W**	1.889
2	0.305	0.919	2.0	2.889
3	0.333	1.337	2.889	4.0

Interesting thing to note here is that (3, 3) is more positive compared to (3, 0) due to the presence of negative reward state at (3, 1). (2, 0) is more positive compared to (1, 1) due to the presence of positive reward at (2, 0). Following the 15th iteration, the utitlities change less than 1%. Hence, the algorithm terminates.