AI HW6

Joshua Wiseman

November 2024

Task 1

For task 1 I used 1000 experiments per state to measure the average return and the following are the results.

State	(1,1)	(1,2)	(1,3)	(2,1)	(2,3)	(3,1)	(3,2)	(3,3)	(4,1)
Utility	-0.132	-0.137	-0.134	-0.116	-0.053	-0.138	-0.144	0.404	-0.221

Table 1: Estimated Utility For Each State via Monte Carlo

Task 2

For task 2 I Implemented the coding solution and got the new policy shown below. You can see the average return of each of the states via the heat map in Figure 1. (Where the direction points towards the highest neighboring value.)

State	(1,1)	(1,2)	(1,3)	(2,1)	(2,3)	(3,1)	(3,2)	(3,3)	(4,1)
Old Policy	R	U	L	U	D	L	L	R	L
New Policy	U	U	R	R	R	U	U	R	L

Table 2: Policy Optimization results with Bellman's Equation

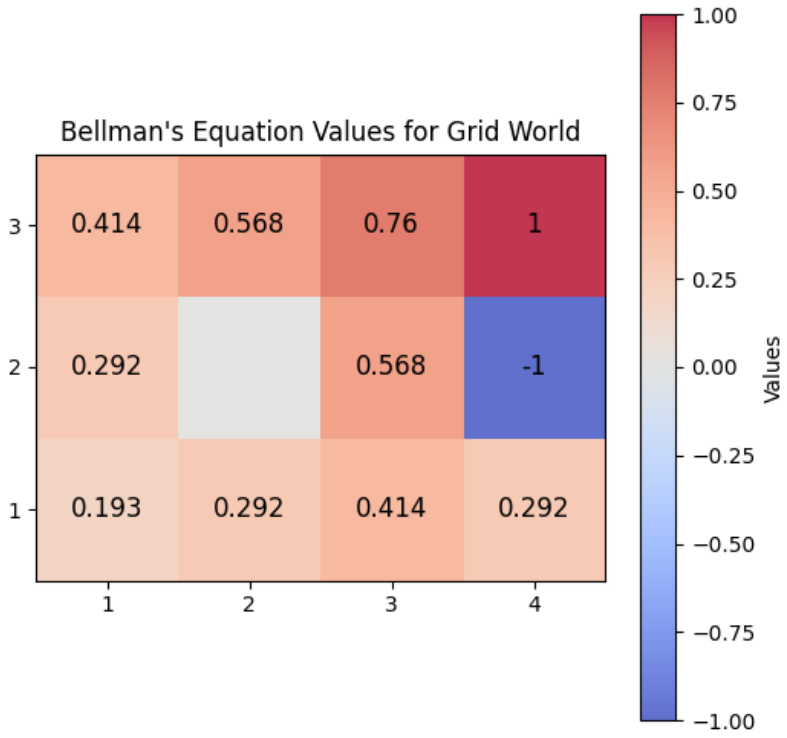

Figure 1: Optimized Policy Heat Map