Imperial College London

COURSEWORK 1

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

COMP97143: Reinforcement Learning

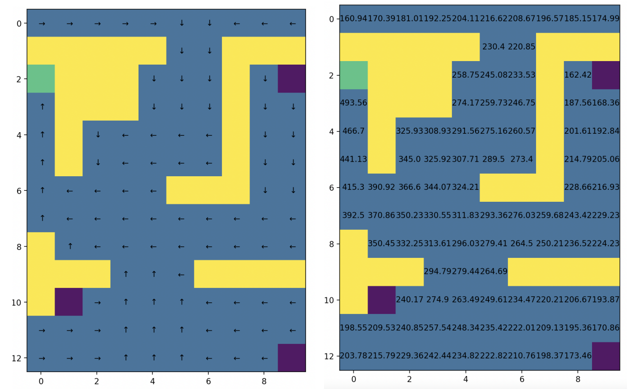
Author:
Jaime Sabal Bérmudez (CID: 01520988)

Date: November 8, 2021

Question 1: Dynamic Programming

1.1: Method Chosen to Solve Grid-World Problem

1.2: Graphical Representation of the Optimal Policy and Value Function



(a) Optimal policy.

(b) Optimal value function.

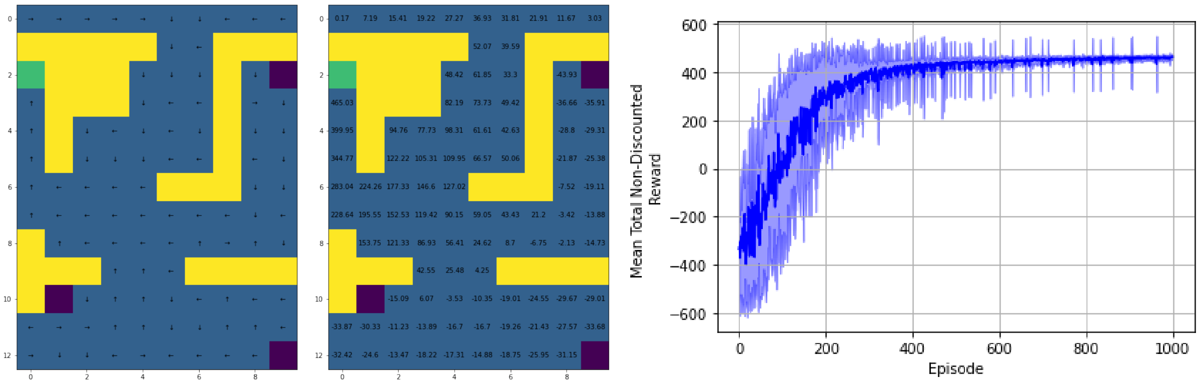
Figure 1: Optimal policy and value function using the CID-personalised parameters $\gamma = 0.96$ and $p = 0.82$.

1.3: The effect of γ and p on the Optimum Policy and Value Function

Question 2: Monte-Carlo Reinforcement Learning

2.1: Method Chosen to Solve Grid-World Problem

2.2-2.4: Graphical Representation of the Optimal Policy and Value Function, and the Learning Curve of Agent

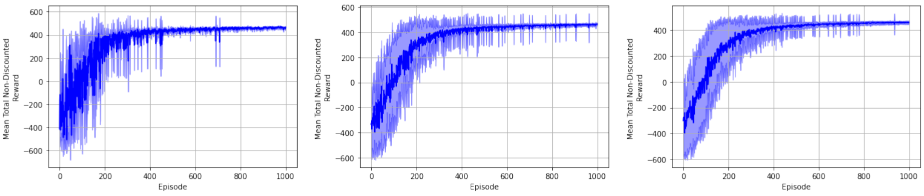


(a) Optimal policy

(b) Optimal value function

(c) Average learning curve

Figure 2: Optimal policy and value function (**a)** and **b)**, respectively); **c)** shows the average learning curve across 25 replications for $1,000$ episodes, (shaded area represents the standard deviation) using a starting epsilon $\epsilon = 0.95$ and a GLIE parameter of 0.999.



(a) 5 **(b)** 25 **(c)** 50

Figure 3: Average learning curves for different numbers of replications on the MC agent: **a)** 5 replications ; **b)** 25 replications ; **c)** 50 replications.

2.5 Effect of ϵ and the GLIE parameter on the Learning Curve of the Agent

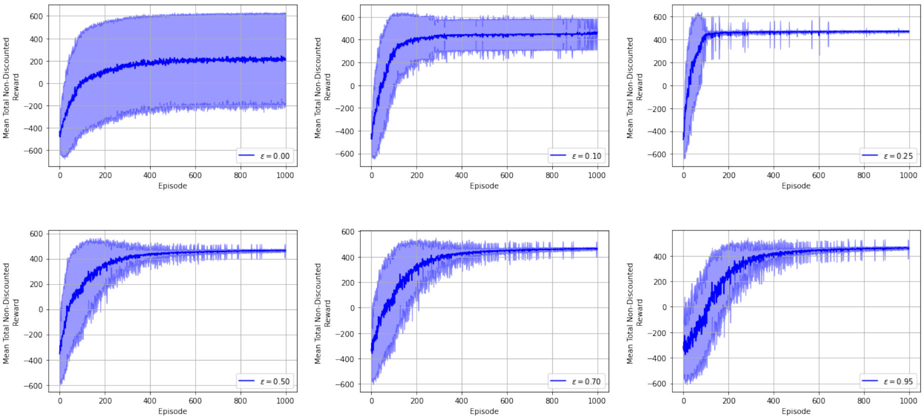


Figure 4: Average total non-discounted rewards across 25 replications for different starting values of epsilon when reducing it by a constant factor of 0.999 (GLIE) after each episode for 1,000 episodes.

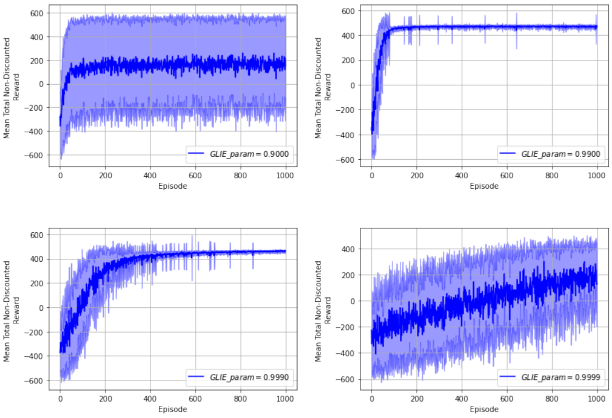
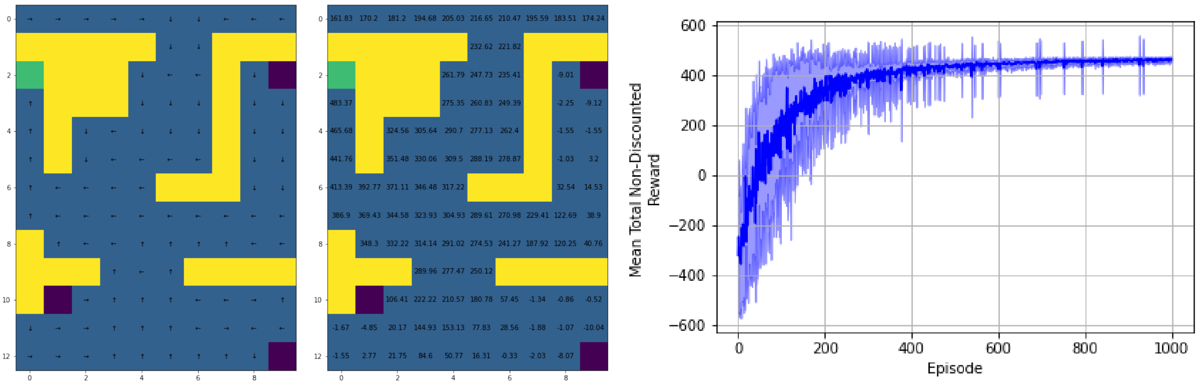


Figure 5: Average total non-discounted rewards across 25 replications for different starting values of the GLIE parameter with a constant $\epsilon = 0.95$.

Question 3: Temporal Difference Reinforcement Learning

3.1: Method Chosen to Solve Grid-World Problem

3.2 - 3.3: Graphical Representation of the Optimal Policy and Value Function, and the Learning Curve of Agent



(a) Optimal policy

(b) Optimal value function

(c) Average learning curve

Figure 6: Optimal policy and value function (**a)** and **b)**, respectively); **c)** shows the average learning curve across 25 replications for $1,000$ episodes, (shaded area represents the standard deviation) using a starting epsilon $\epsilon = 0.95$ and a GLIE parameter of 0.999.

3.4: Effect of ϵ and α on the Learning Curves of the Agent

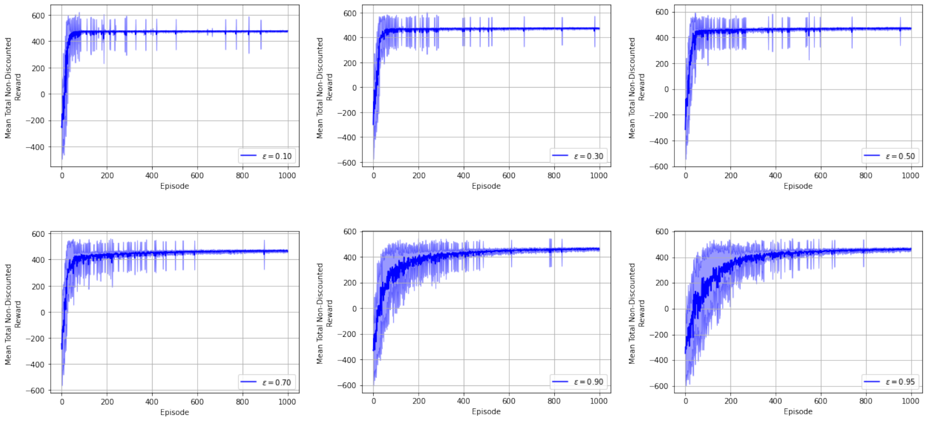


Figure 7: Average total non-discounted rewards across 25 replications for different starting values of epsilon when reducing it by a constant factor of 0.999 (GLIE) after each episode for 1,000 episodes.

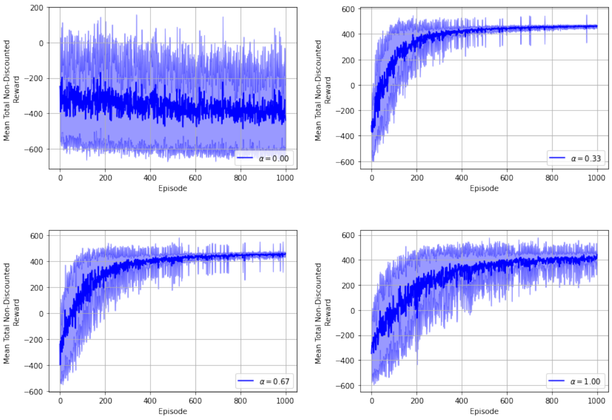


Figure 8: Average total non-discounted rewards across 25 replications for different learning rates α with a constant $\epsilon = 0.95$.

Question 4: Comparison of Learners

4.1-4.2: Value Function Estimation Error for MC and TD Learners

4.3

4.4