Analysis of SARSA, Q-Learning and Monte-Carlo Techniques on Taxi Environment

Course Project - Introduction to Reinforcement Learning

Laiba Jamil
Muhammad Zain Yousuf
Hamad Abdul Razzaq

Problem
Taxi Environment

- 2D Grid → 5 × 5
- Driver needs to pick and drop the passenger
- Passenger can be in 4 locations
- Destination can also be 4 locations
- We will model this problem as MDP

Modeling the Problem as MDP

- **State Space**
  - 4-tuple

$$s = \left(d_x, d_y, p_p, p_d\right)$$

  - d_x: x-position of the driver
  - d_y: y-position of the driver
  - p_p: Current position of Passenger
  - p_d: Drop Off Location

Modeling the Problem as MDP

- **Action Space**
  - Driver can move in 4 directions
  - Driver can pick/drop the passenger

$$A = \{\text{ Left, Right, Up, Down, Pick, Drop }\}$$

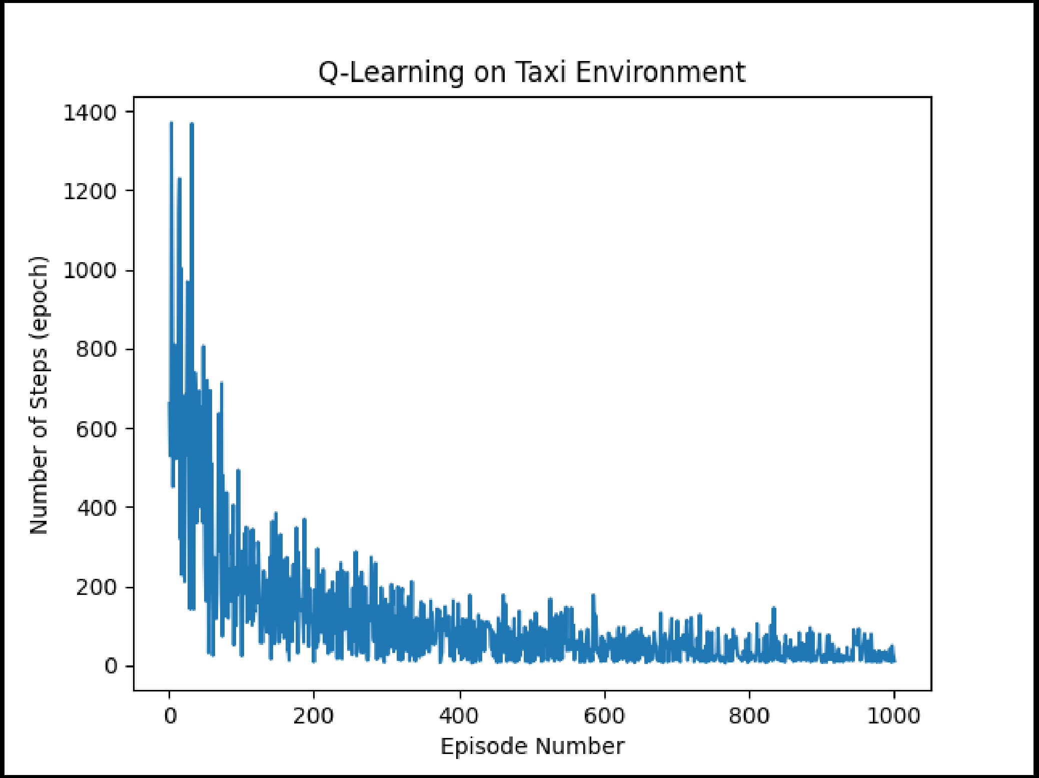
Modeling the Problem as MDP

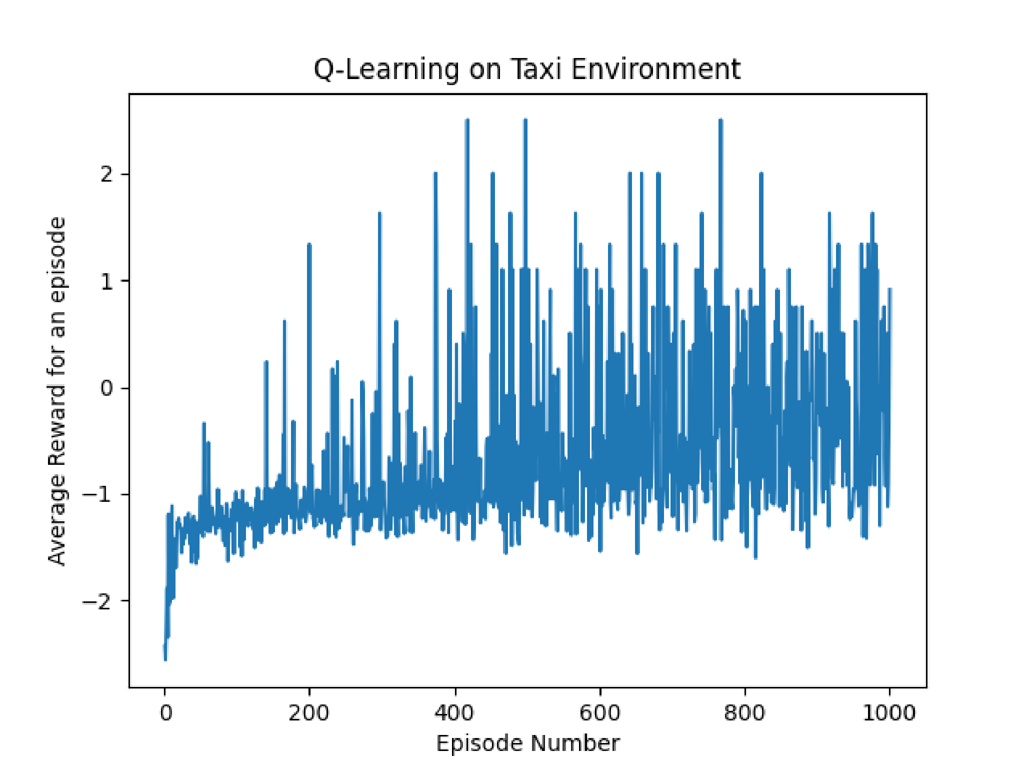
- **Reward Shaping**
    - Driver moves in either direction → $r = -1$
    - Driver drops passenger in wrong location → $r = -10$
    - Driver drops passenger in correct location → $r = 20$

Hyper-Parameters

- $N = 1000$ Episodes
- Max Episode Length = 50000
- $\gamma = 0.7$ (Discount Factor)
- $\alpha = 0.1$ (Learning Rate)
- $\epsilon = 0.1$ (Exploration)

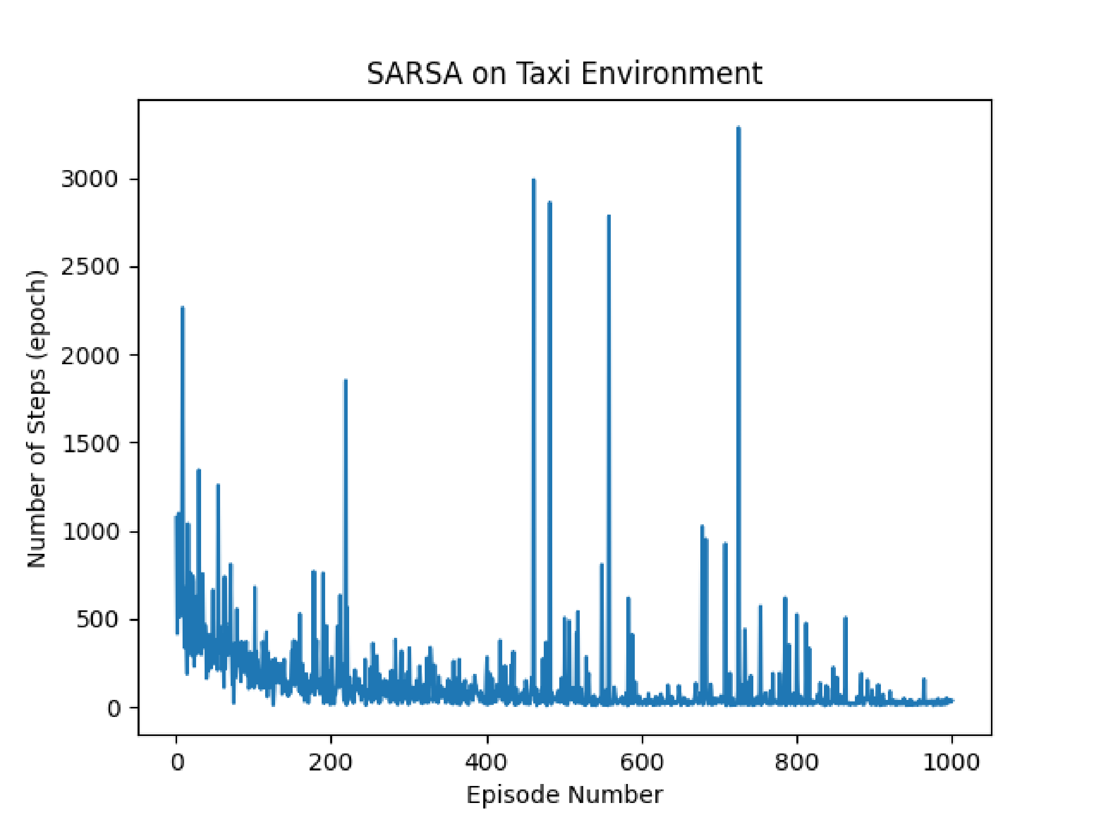
Applying Q-Learning Technique

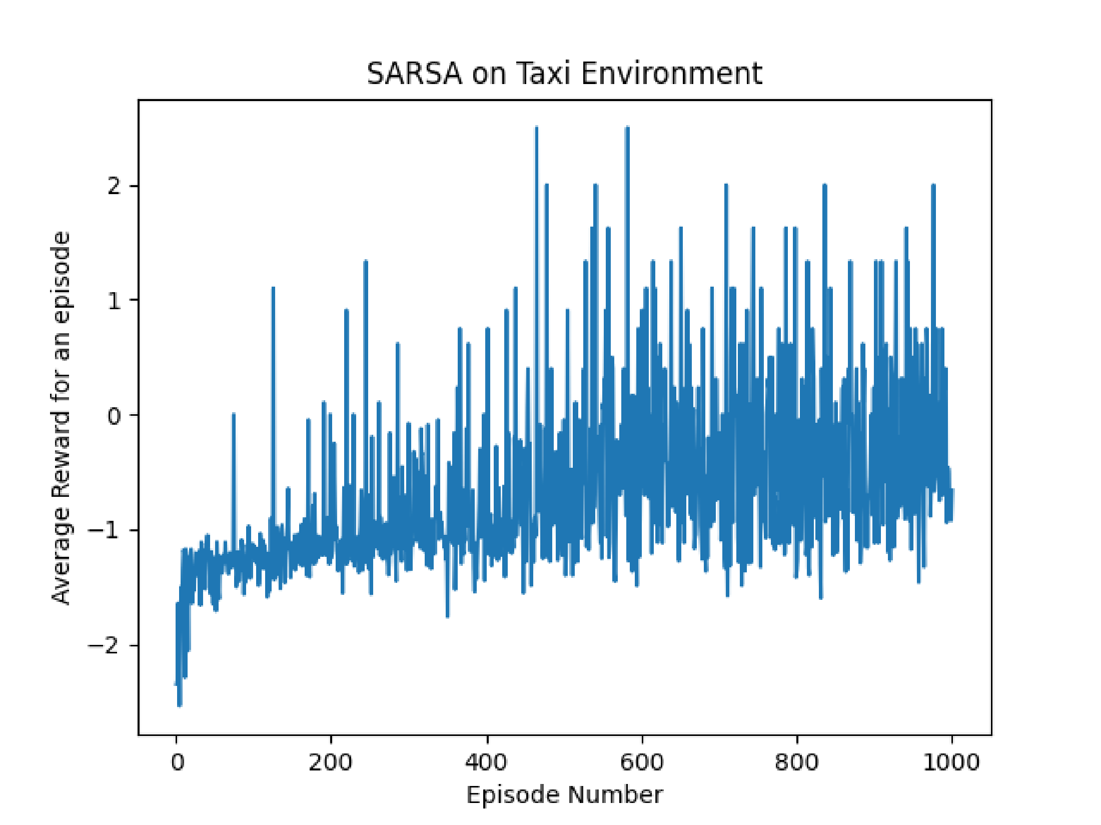




Average Reward = 0.13

Applying SARSA Technique

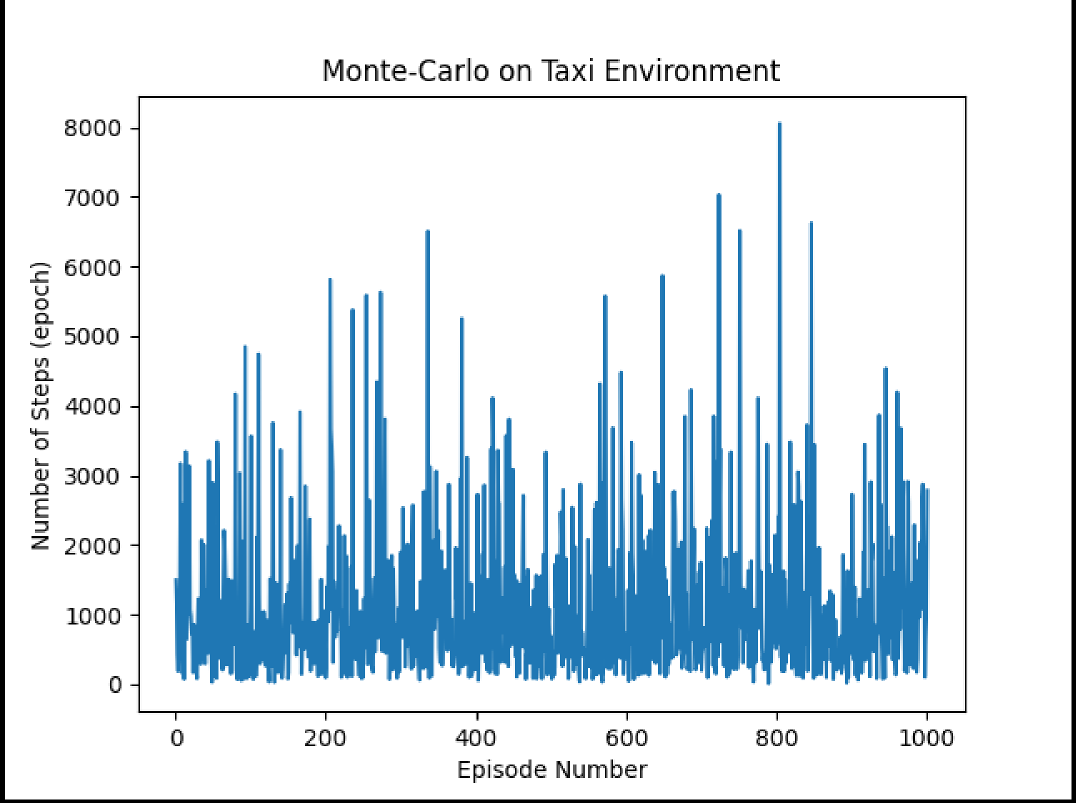


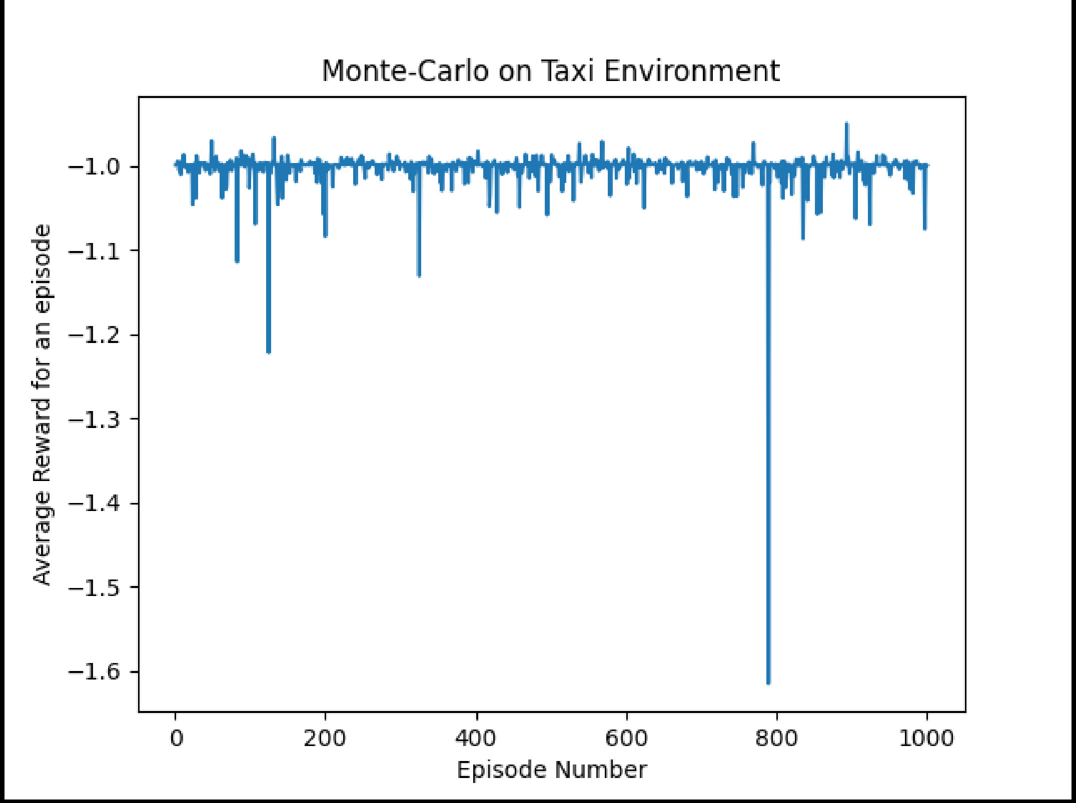


Average Reward = −0.09

Applying Monte-Carlo Technique

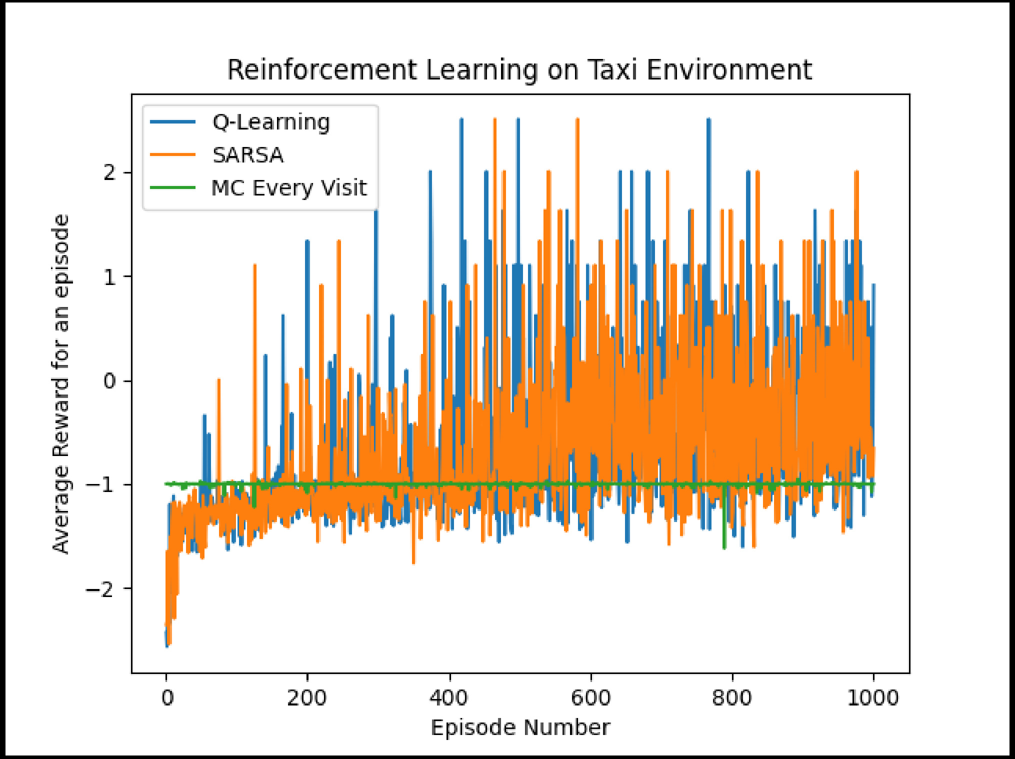
- Every Visit

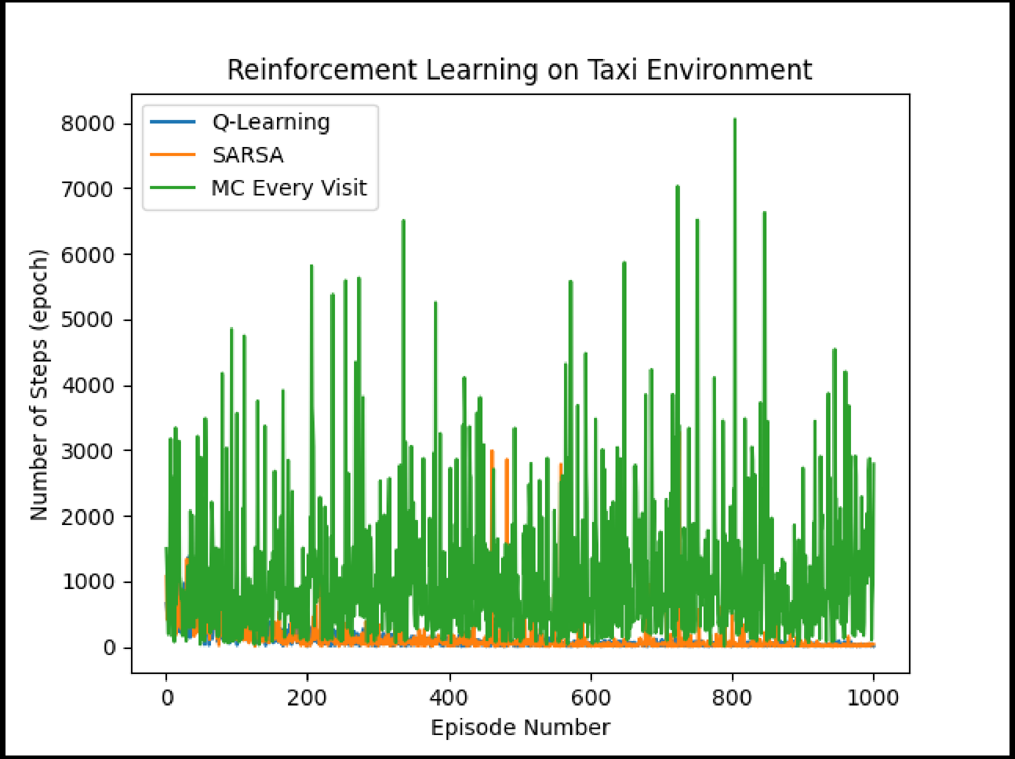




$$\boxed{\text{Average Reward} = -1.002}$$

Analysis / Comparison





Q-Learning Outperforms!

Thank you!

References:

- [1] https://towardsdatascience.com/solving-the-taxi-environment-with-q-learning-a-tutorial-c76c22fc5d8f
- [2] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction. Cambridge, MA: The MIT Press, 2020.