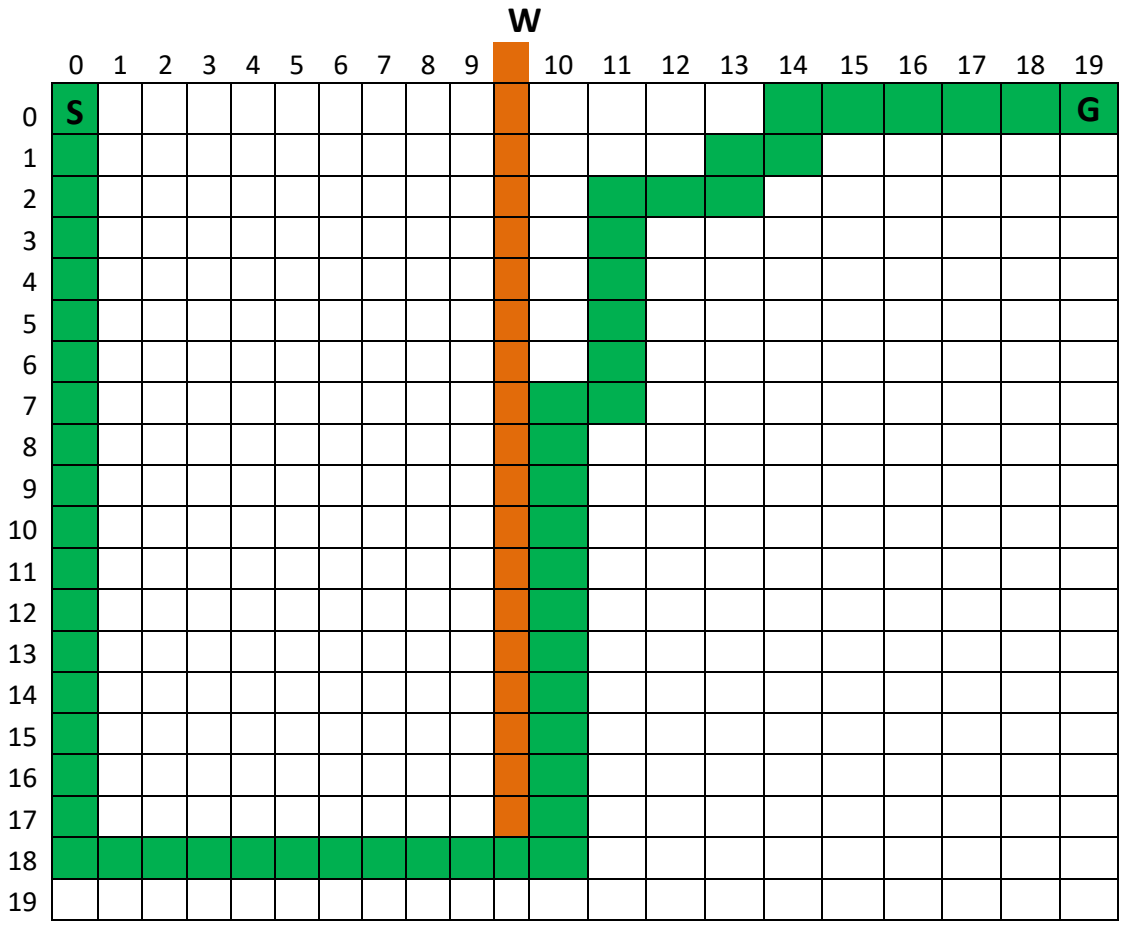
Q-Learning

Mohammad Mirzanejad

- **Optimal policy with a length of 56**



W= Wall

S=Start point

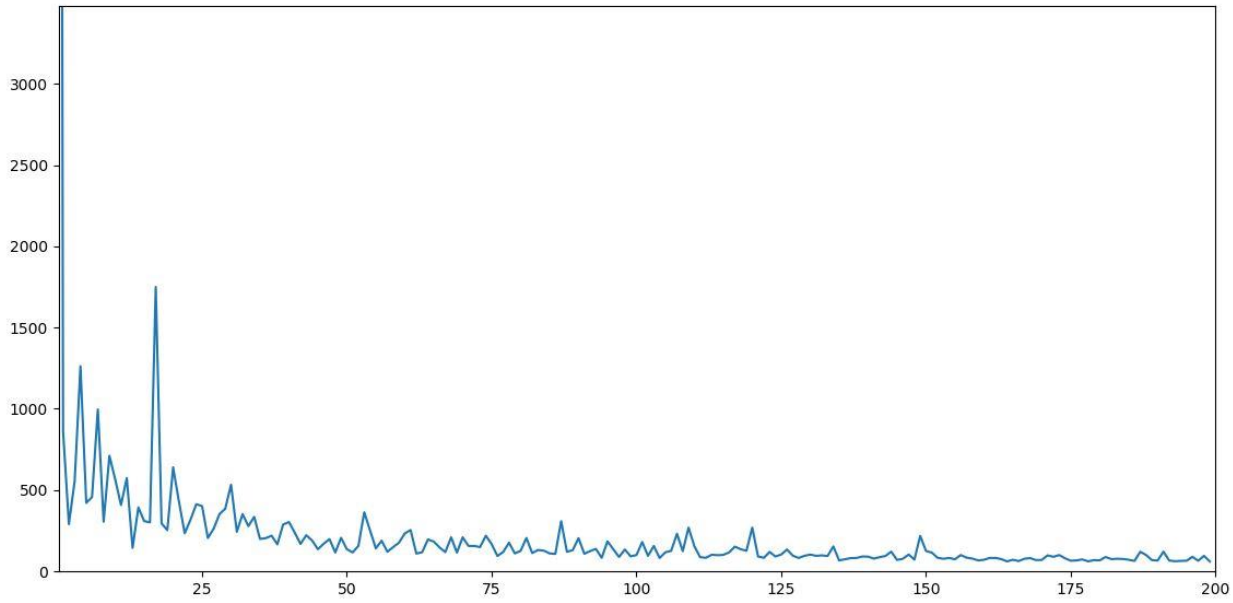
G= Goal

Selected path:

[(0, 0), (1, 0), (2, 0), (3, 0), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (3, 7), (3, 8), (4, 8), (5, 8), (6, 8), (6, 9), (7, 9), (8, 9), (9, 9), (10, 9), (11, 9), (12, 9), (13, 9), (14, 9), (15, 9) , (16, 9), (17, 9), (18, 9), (18, 10), (17, 10), (16, 10), (15, 10), (14, 10), (13, 10), (13, 11), (12, 11), (11, 11), (10, 11), (10, 12), (10, 13), (9, 13), (8, 13), (7, 13), (6, 13), (5, 13), (4, 13), (4, 14), (3, 14), (3, 15), (2, 15) , (1, 15), (0, 15), (0, 16), (0, 17), (0, 18), (0, 19)]

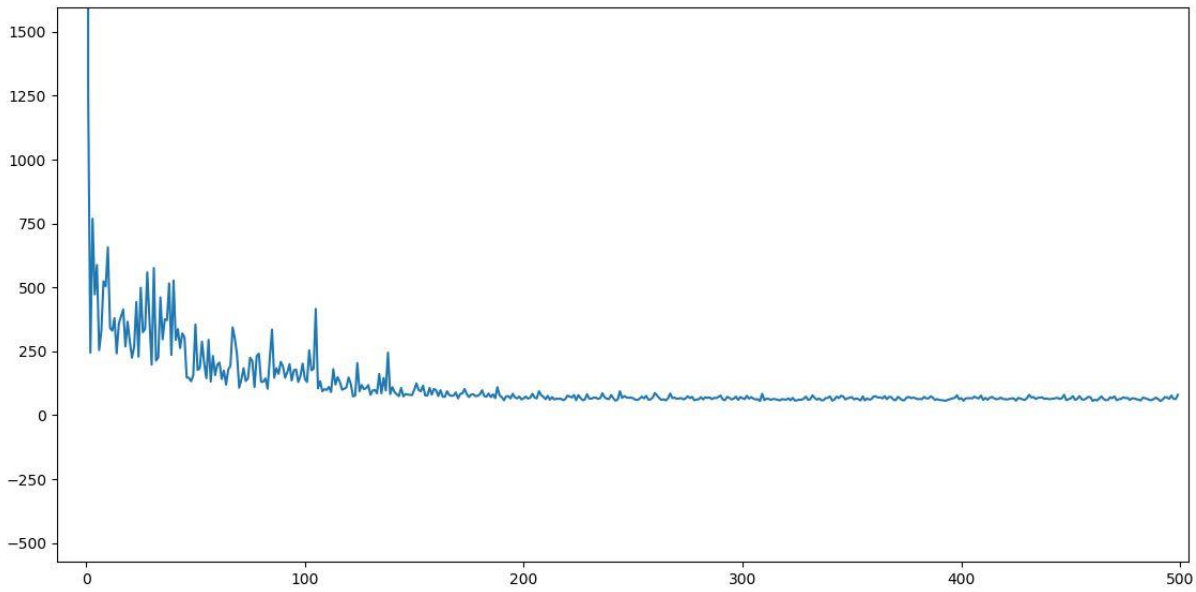
Len: 56

the agent's experience



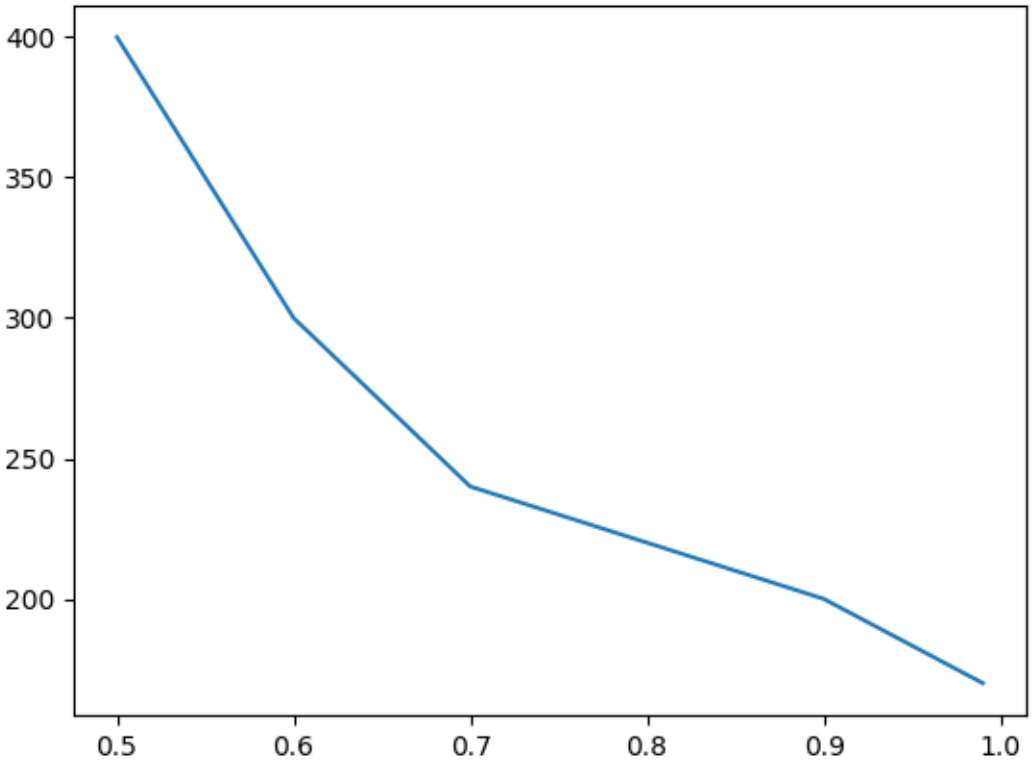
Q table are updated accordingly, and the agent can make more optimal choices and take fewer steps towards the goal, but in the meantime, the overall trend of the number of steps in each episode is decreasing. As can be seen in the diagram, there are fluctuations, that is, the number of steps in the episode goes up and down, which is due to the 15% probability considered for choosing random moves other than the optimal and higher Q value move in order to comply with the principle of exploration. All the paths in the table are considered in the

training phase. This graph is drawn for a learning rate of 0.8 and 200 repetitions, which resulted in the generation of an optimal path with a length of Of course, in the case of multiple executions, due to the possibility of randomly choosing the path, we may reach convergence for a higher number of iterations (about more than 10 or 20) or the optimal path length may be more than 56 which was observed in the performed implementations.



- **:The effect of different gamma values on algorithm convergence**

  The closer the value of gamma is to zero, the number of episodes required for convergence increases. In fact, the agent focuses more on immediate rewards, and the closer the gamma value is to one, the agent assigns more weight todelayed rewards. Below are examples of different gamma values andthe number of episodes.



Convergence	Fewest moves in an episode	Number of episodes	gamma
no	103	100	0.99
no	75	150	0.99
Yes	60	170	0.99
no	59	170	0.9
Yes	58	200	0.9
Yes	59	220	0.8
Yes	58	240	0.7
no	57	260	0.6
Yes	55	300	0.6
no	57	380	0.5
Yes	55	400	0.5
no	103	500	0.3