CS533 Implementation Homework 3

Ernst Adrian Henle

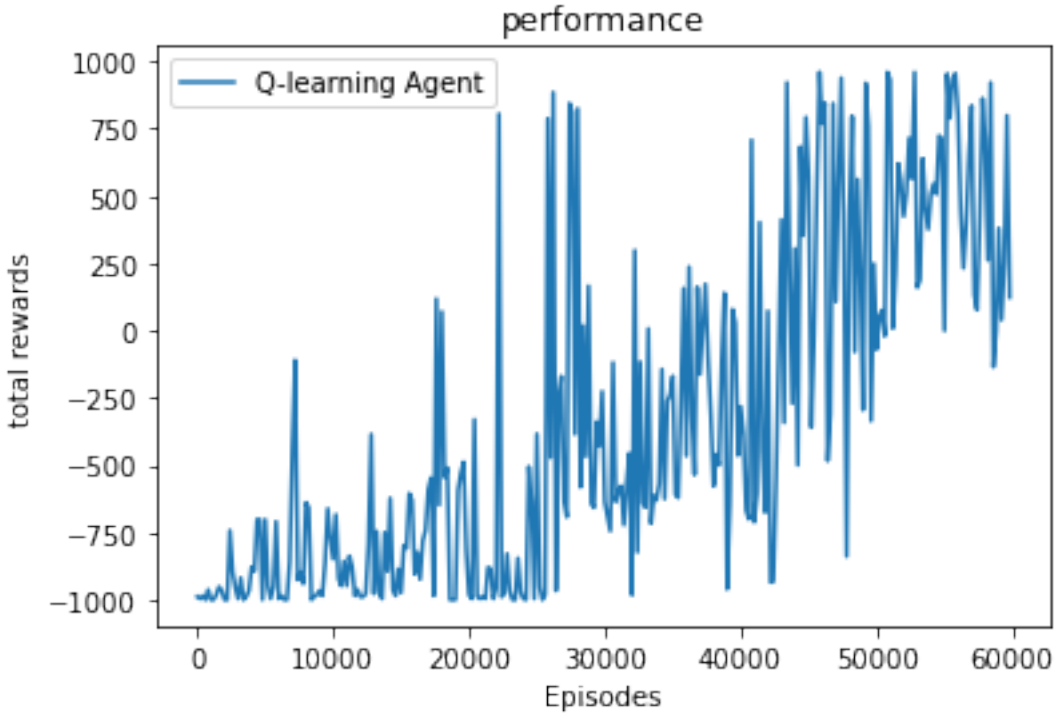
12 May 2021

1 Learning Curves

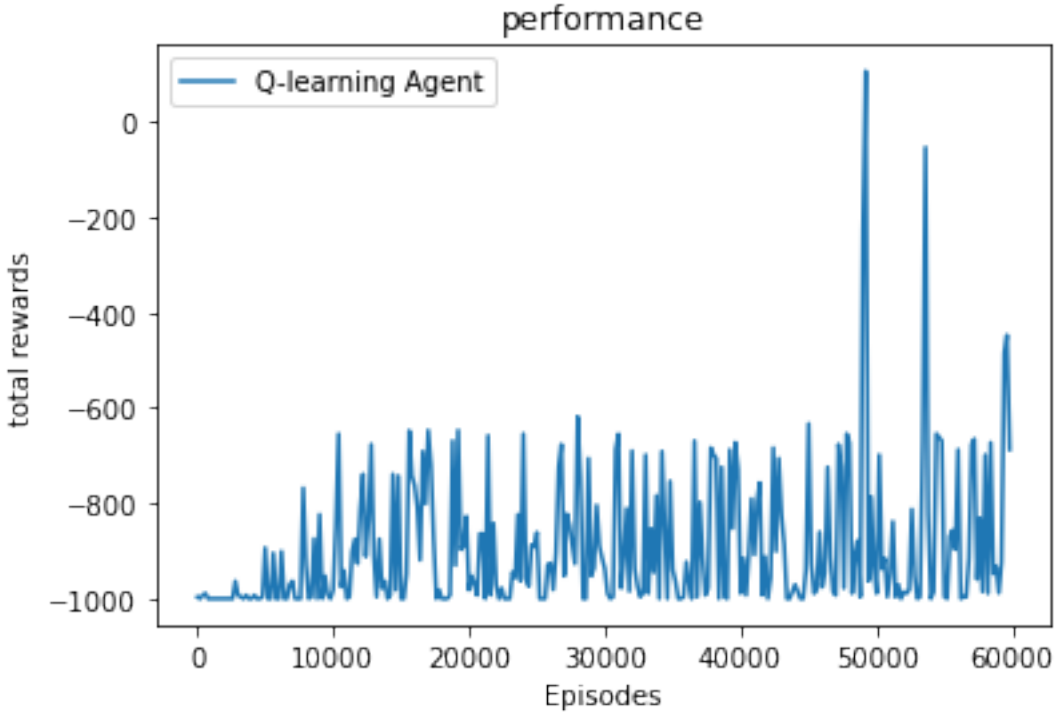
1.1 Q-Learning Agent, Dangerous Hallway

Tests were performed with 60,000 learning episodes and a test interval of 200 episodes.

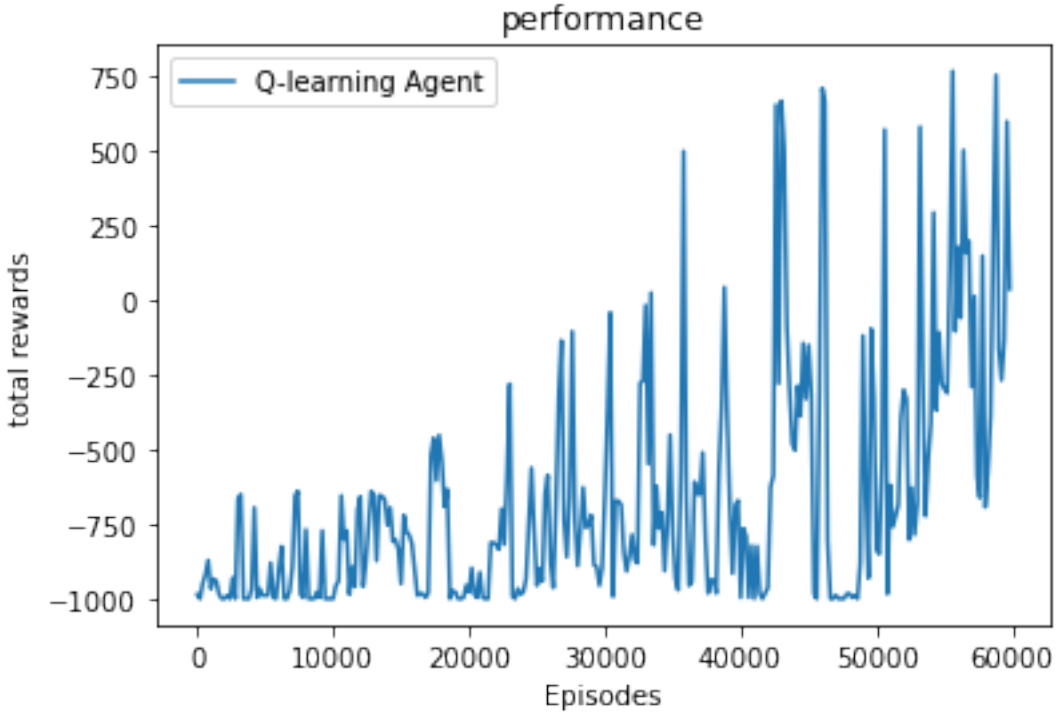
1.1.1 $\epsilon = 0.1$, $\alpha = 0.1$



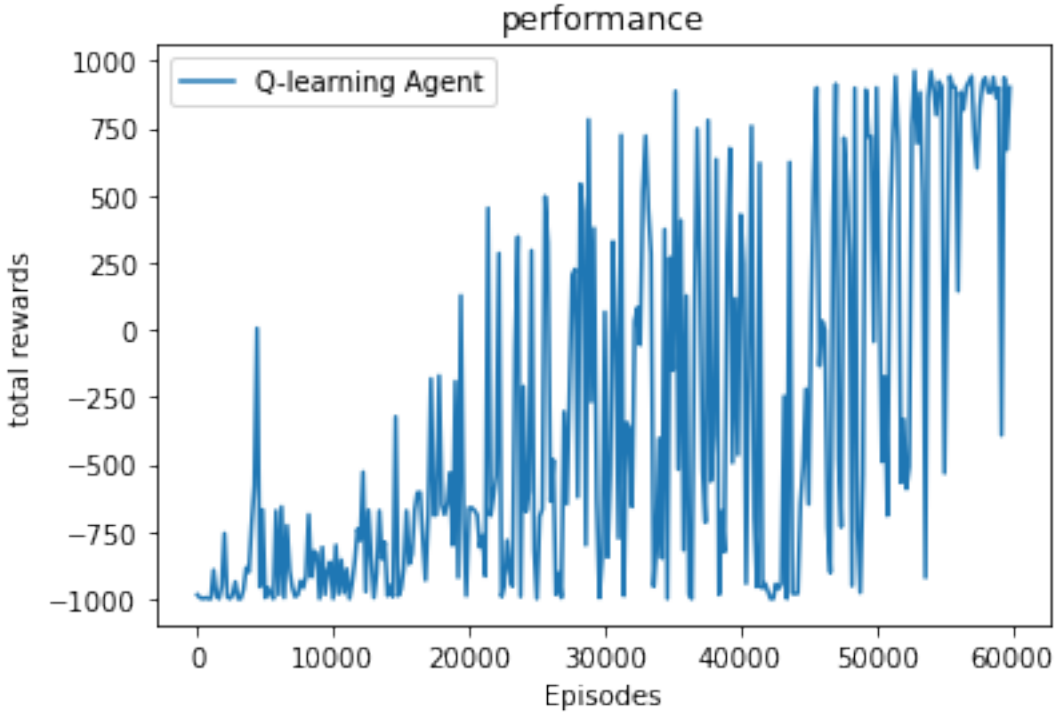
1.1.2 $\epsilon = 0.1$, $\alpha = 0.001$



1.1.3 $\epsilon = 0.3$, $\alpha = 0.1$



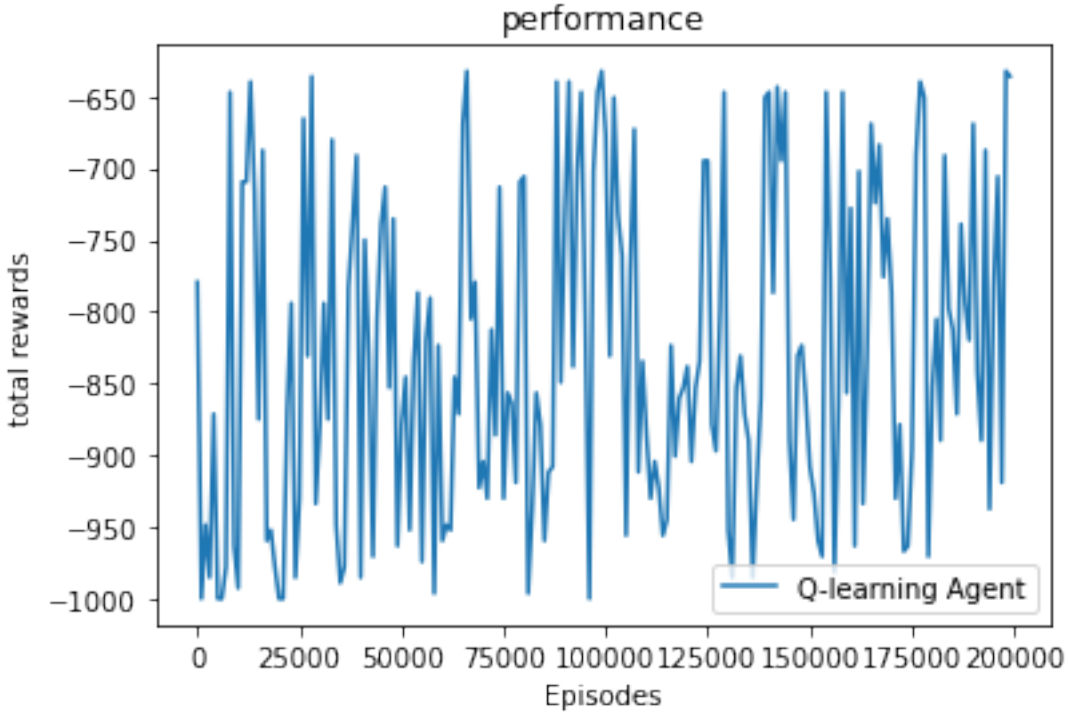
1.1.4 $\epsilon = 0.005$, $\alpha = 0.1$



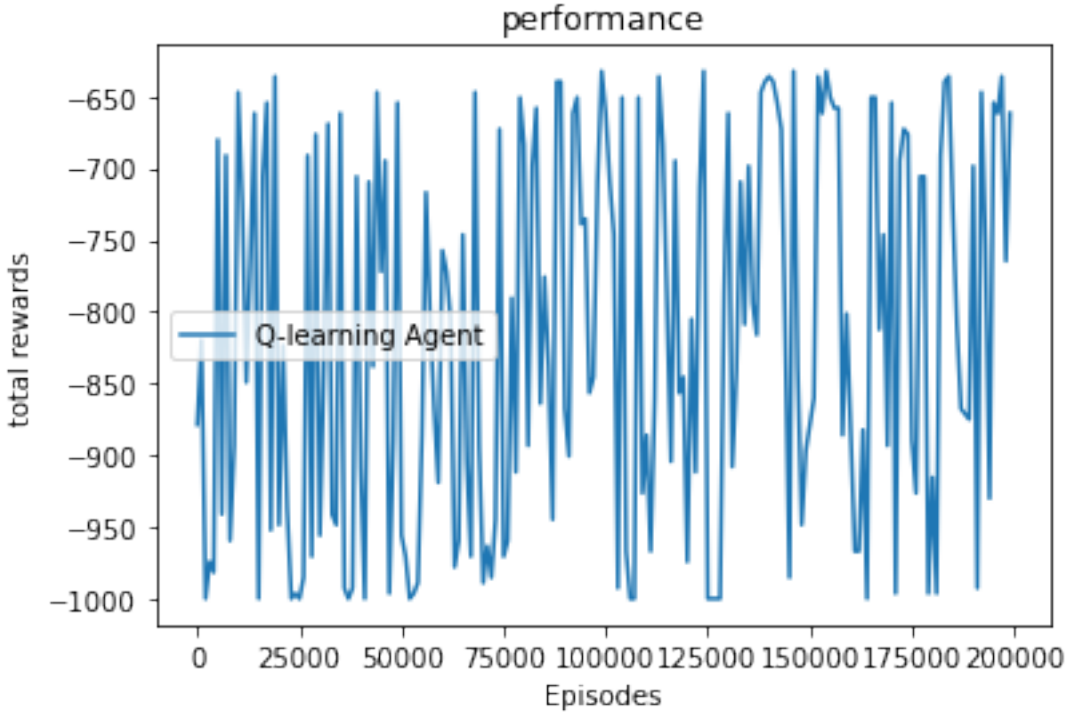
1.2 Q-Learning Agent, 16x16

Tests were performed with 200,000 learning episodes and a test interval of 1,000 learning episodes.

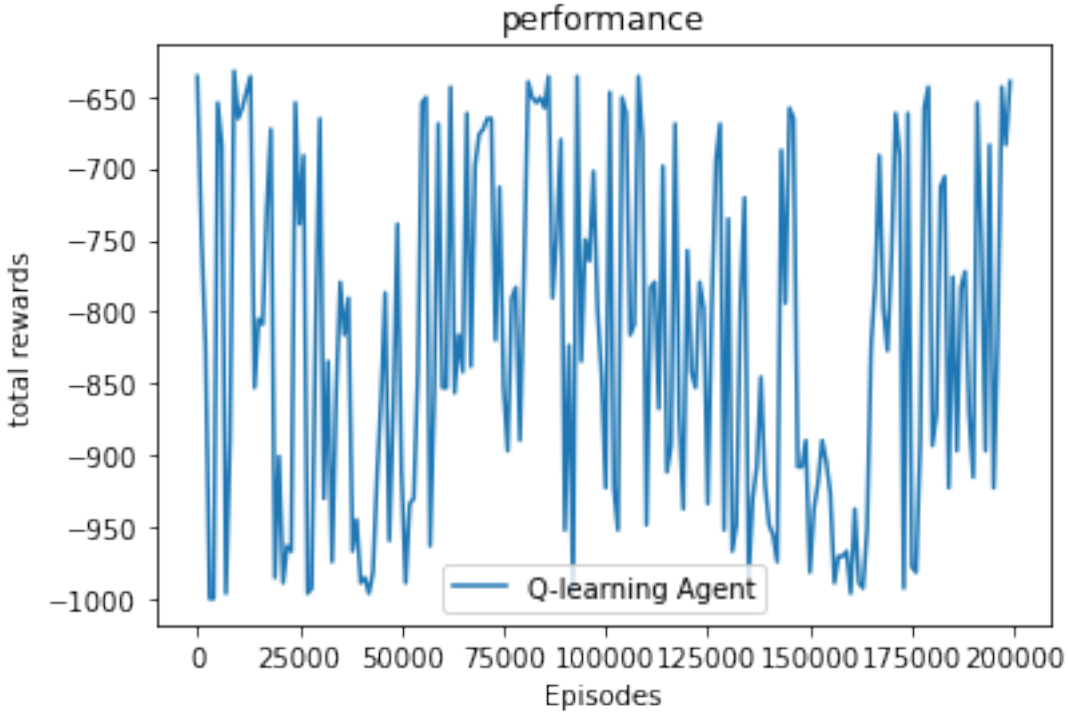
1.2.1 $\epsilon = 0.1$, $\alpha = 0.1$



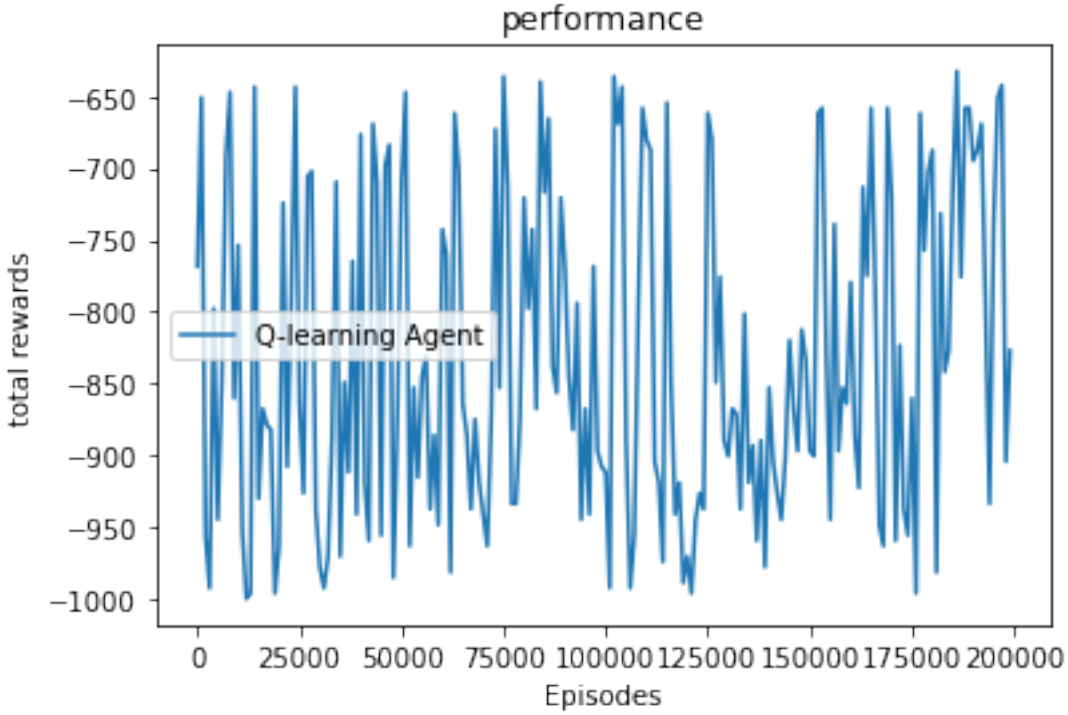
1.2.2 $\epsilon = 0.1$, $\alpha = 0.001$



1.2.3 $\epsilon = 0.3$, $\alpha = 0.1$



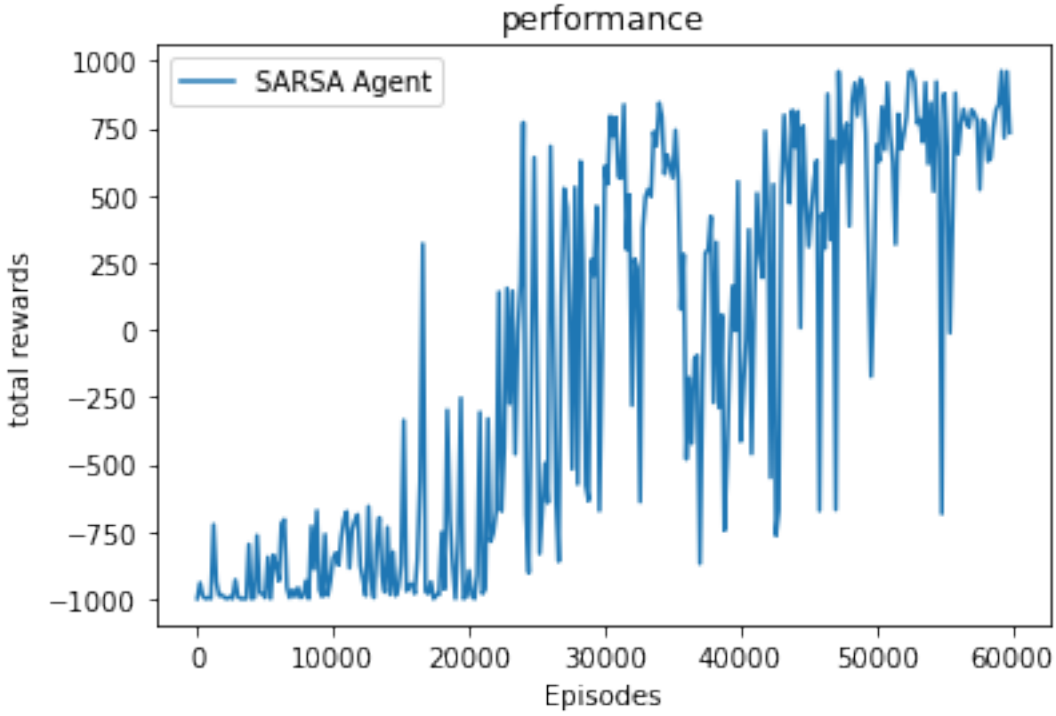
1.2.4 $\epsilon = 0.005$, $\alpha = 0.1$



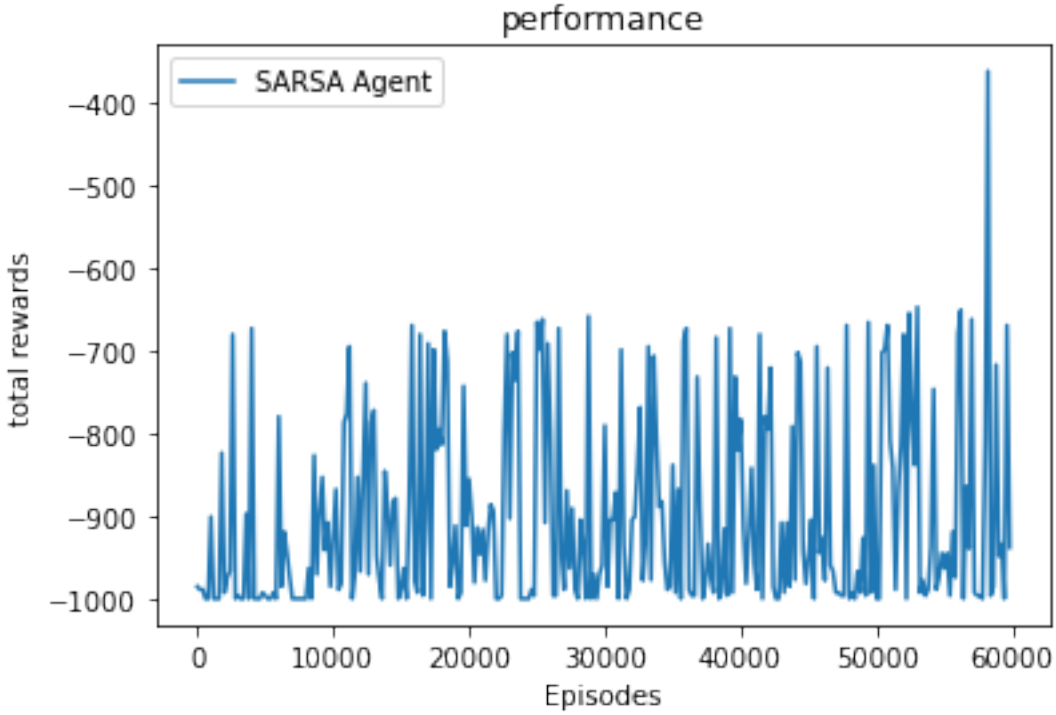
1.3 SARSA Agent, Dangerous Hallway

Tests were performed with 200,000 learning episodes and a test interval of 1,000 episodes.

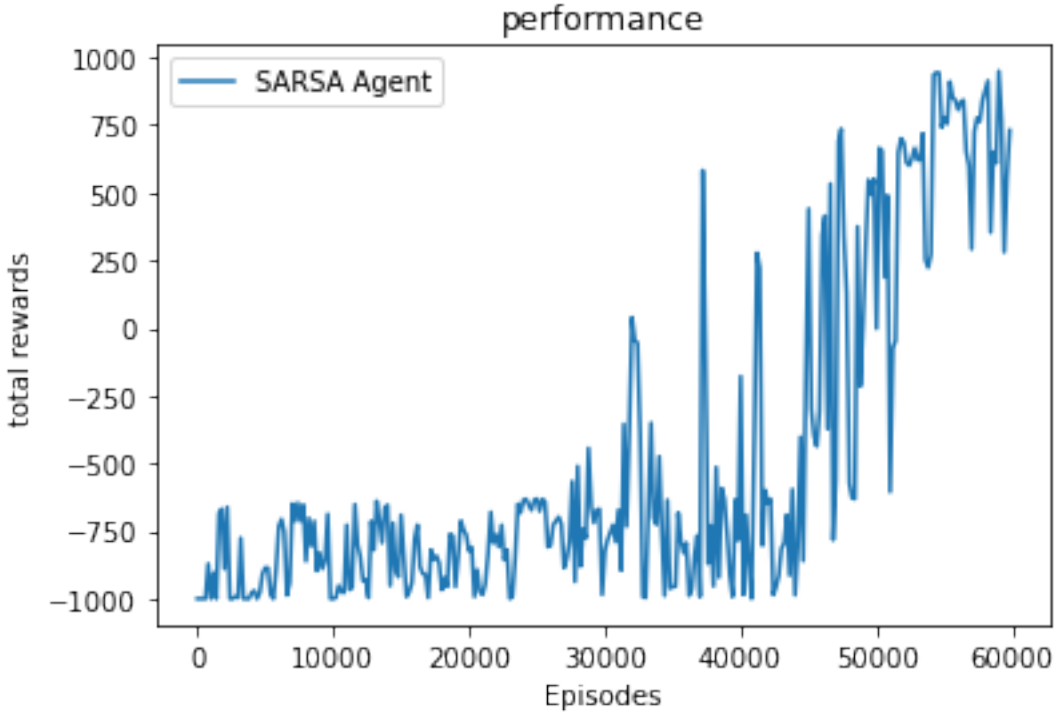
1.3.1 $\epsilon = 0.1$, $\alpha = 0.1$



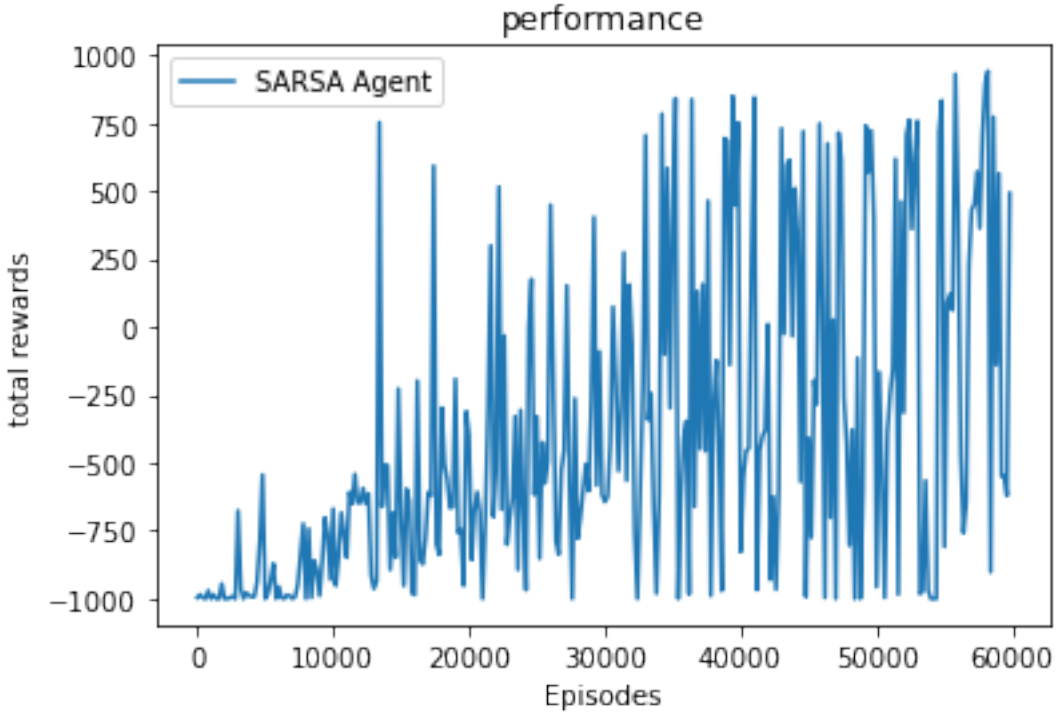
1.3.2 $\epsilon = 0.1$, $\alpha = 0.001$



1.3.3 $\epsilon = 0.3$, $\alpha = 0.1$



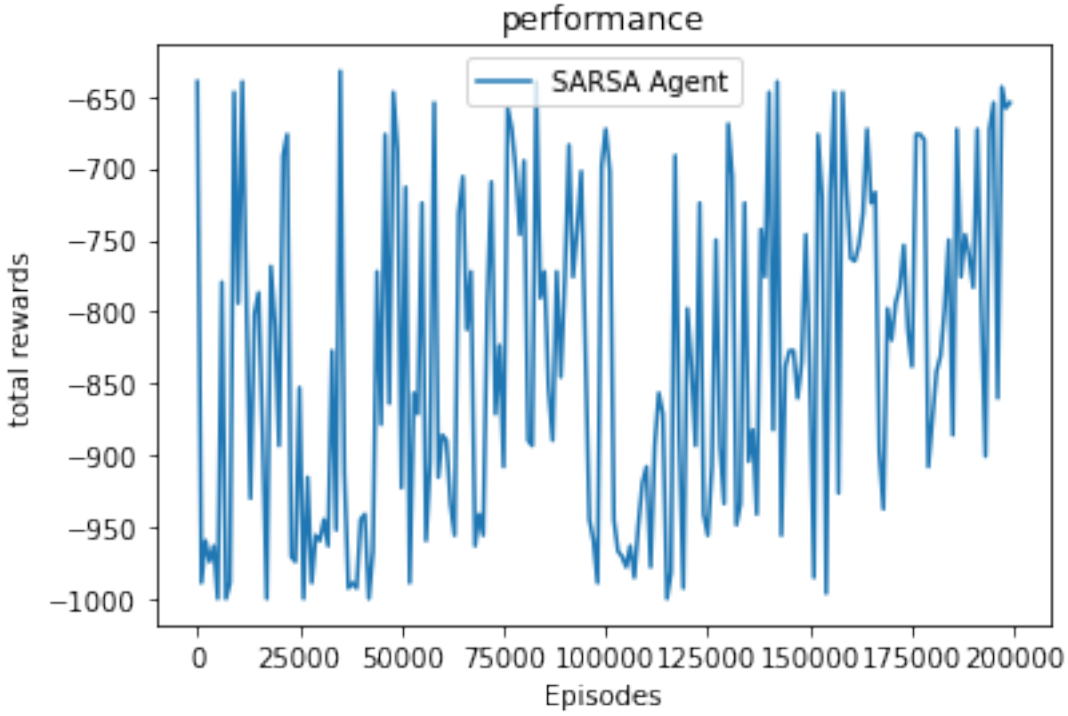
1.3.4 $\epsilon = 0.005$, $\alpha = 0.1$



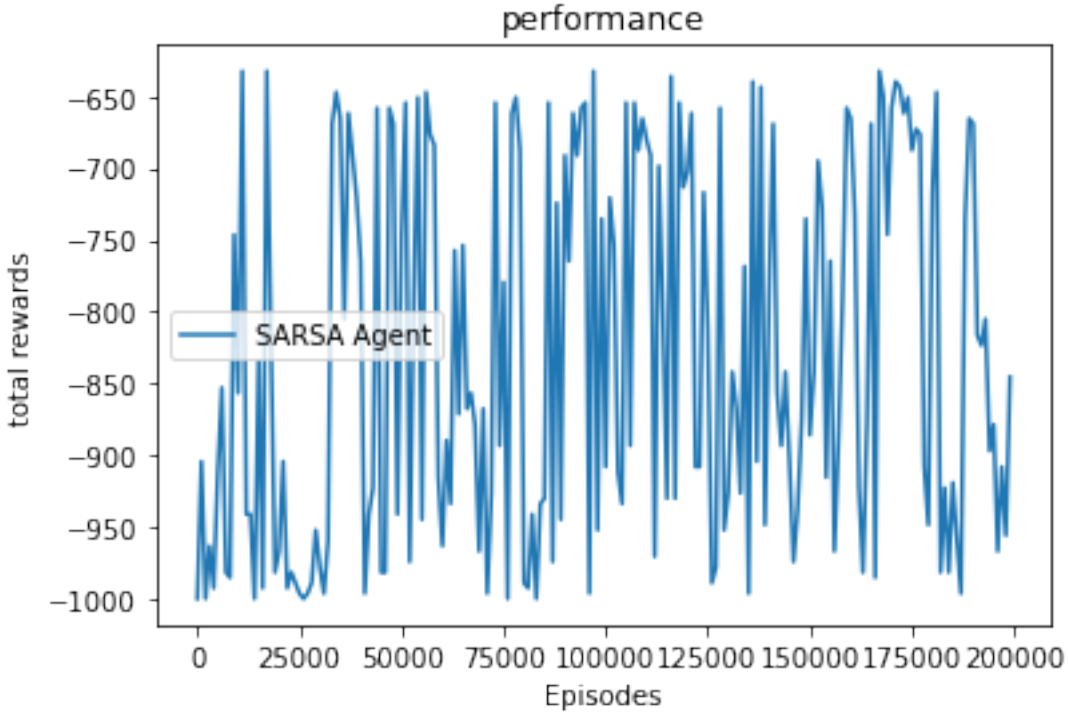
1.4 SARSA Agent, 16x16

Tests were performed with 200,000 learning episodes and a test interval of 1,000 episodes.

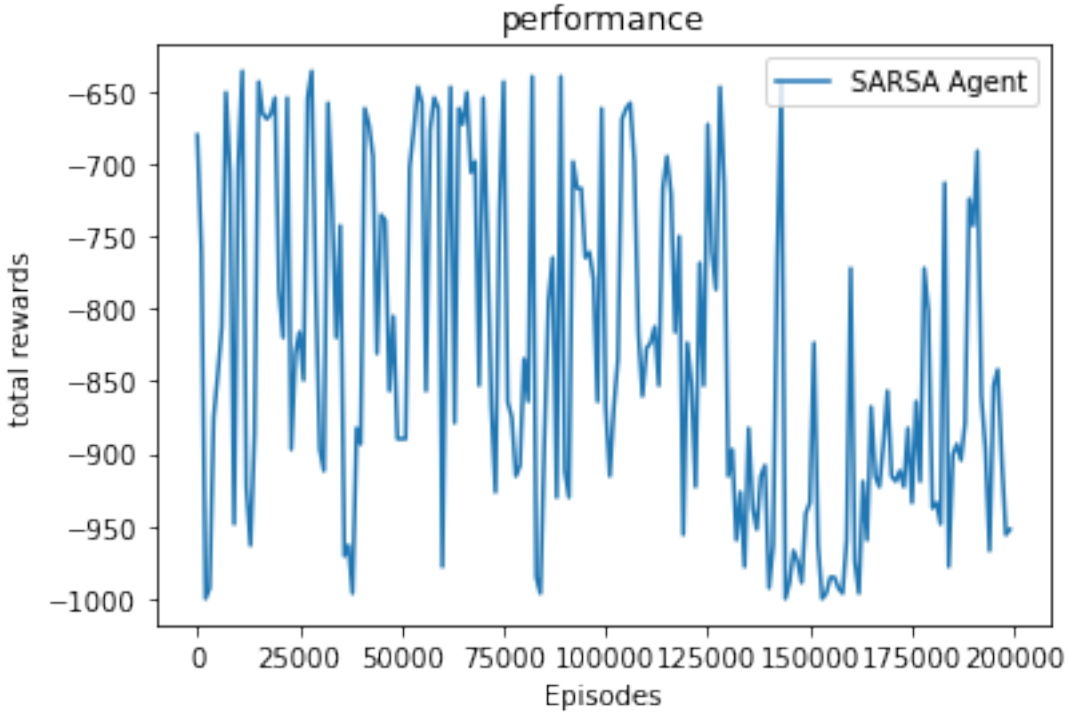
1.4.1 $\epsilon = 0.1$, $\alpha = 0.1$



1.4.2 $\epsilon = 0.1$, $\alpha = 0.001$



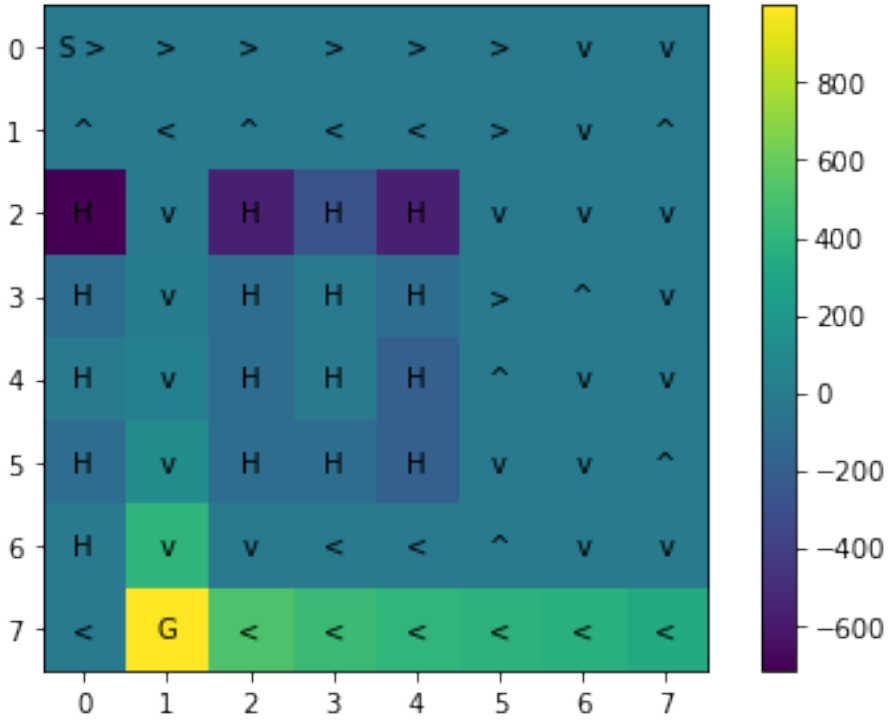
1.4.3 $\epsilon = 0.3$, $\alpha = 0.1$



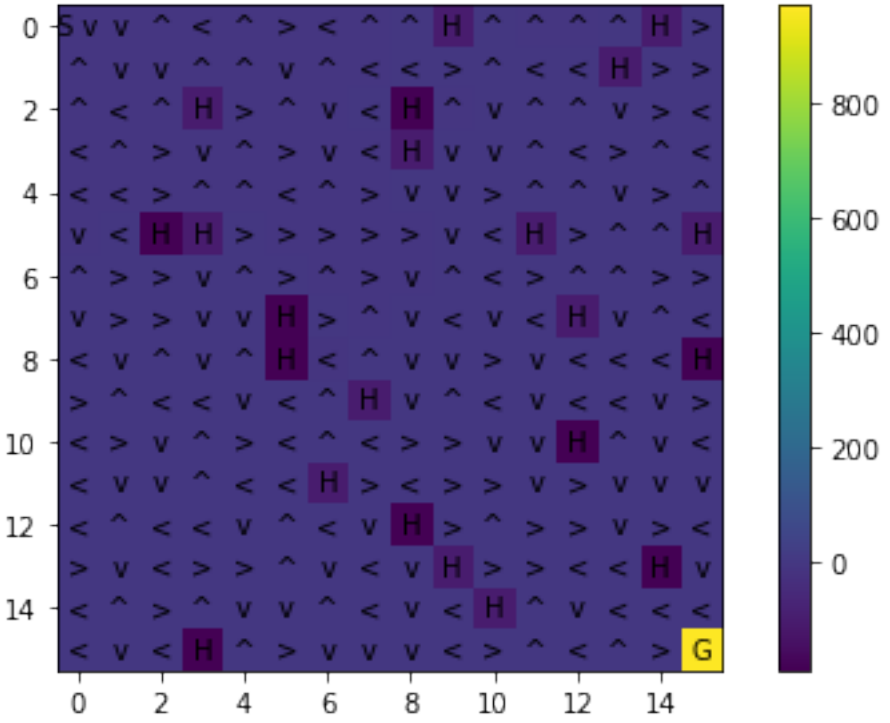
1.4.4 $\epsilon = 0.005$, $\alpha = 0.1$

2 SARSA: Different Learning Rates and Exploration Factors

For the Dangerous Hallway map, the SARSA Agent performed best with a learning rate $\alpha = 0.1$ and exploration factor $\epsilon = 0.3$

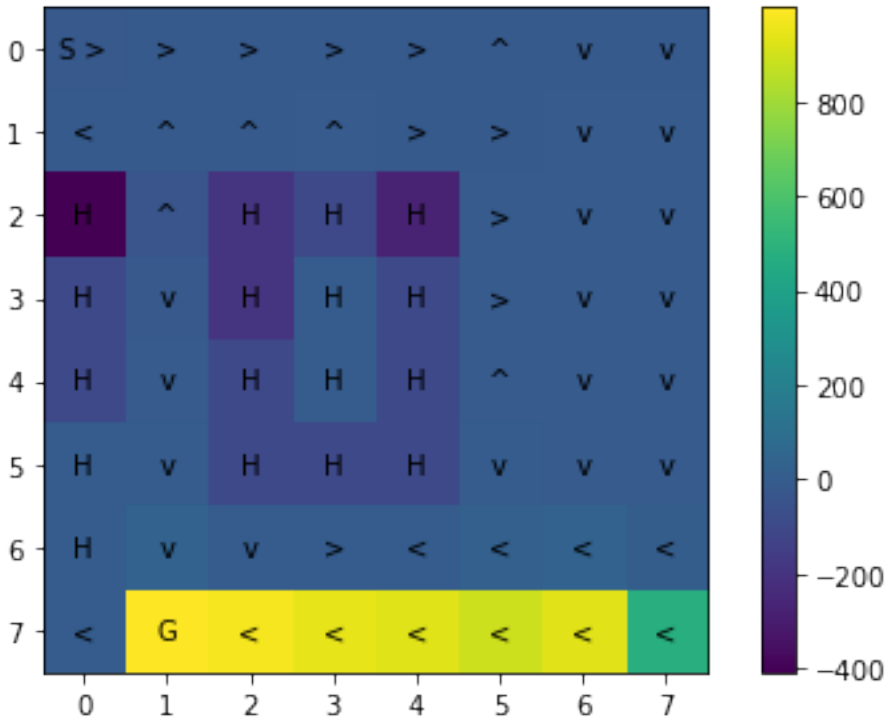


For the 16x16 map, the SARSA Agent performed best with a learning rate $\alpha = 0.1$ and exploration factor $\epsilon = 0.005$

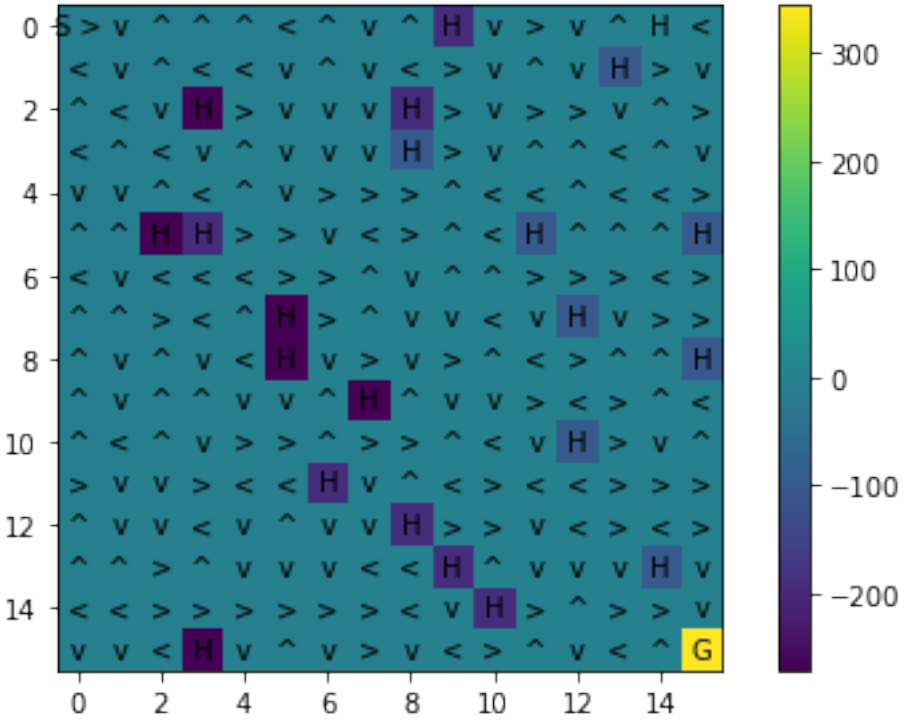


3 Q-Learning: Different Learning Rates and Exploration Factors

For the Dangerous Hallway map, the Q-Learning Agent performed best with a learning rate $\alpha = 0.1$ and exploration factor $\epsilon = 0.1$



For the 16x16 map, the Q-Learning Agent performed best with a learning rate $\alpha = 0.1$ and exploration factor $\epsilon = 0.1$



4 Differences Between Q-Learning and SARSA

SARSA appeared to learn the Dangerous Hallway map somewhat better than Q-Learning. In particular, the policy calls for moving directly away from the lower edge of the large central hole, into the "home stretch" region which directs the agent into the goal.

5 Distributed Implementation

I was unable to run the distributed tests to completion, due to the algorithm seeming to hang indefinitely. With a correct implementation, I would expect the policies learned to be similar, though not necessarily identical, and for the processes to converge on good policies in less run-time for the distributed implementation. The speed should increase roughly proportionally with the number of workers. Eliminating the interleaved evaluation will greatly accelerate the program.