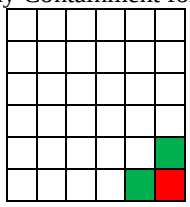
CSE 410: Final Project Milestone

Peter M. VanNostrand

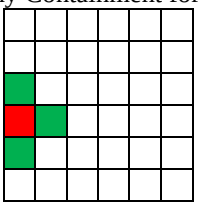
1 Description

As described in our final project proposal, the goal of this exercise is to train a set of Deep Q-Learning agents to collaborate and "contain" and enemy. This is done by positioning an agent on each side of the enemy so that it would be unable to move. Samples of this are shown for $n = 2,3,4$ agents below on square gridworld of size $s = 5$, with the agents shown in green and the enemy shown in red..

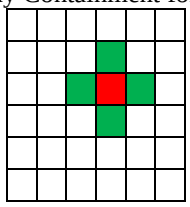
Enemy Containment for $n = 2$



Enemy Containment for $n = 3$



Enemy Containment for $n = 4$



2 Milestone

The milestone for this project is solving the $n = 2$ case with simple Q-Learning. The environment for this project is modified from our assignment one submission, and consists of a deterministic grid world built to comply with the OpenAI gym environment specification. Modifying this environment to accept multiple agents rather than just one, was straightforward, but tedious. Essentially, we tracked down every instance where the environment interacted with the agent, and vectorized the implementation. This consisted of converting previous instance variables to vectors, such as the location of the agent and the starting position of the agent, as well as modifying operations on the agent to be iterable. For example, when the environment steps an action, it now receives a list of actions which correspond to the list of agents, the environment was modified to move each agent independently and check for collisions of agents with each other, with edges of the grid, and with the enemy. Agents are acted on by their ordering, with preference for movement given to agents which appear earlier in the environments list.

Once the environment was modified to support multiple agents, the training algorithm was similarly adjusted to handle Q-Learning of multiple agents. Conveniently the agents themselves did not need to be modified for this process, as they are substantially the same. Substituting these Q-Learning agents for the DQ-Learning agents developed in assignment two should be straightforward as we move forward in the project.

3 Results

Below are shown the results of training two Q-Learning agents on a gridworld of size 4x4. Hyperparameters were tuned to ensure the minimum possible training time for this configuration, and achieve convergence in less than 50 iterations.

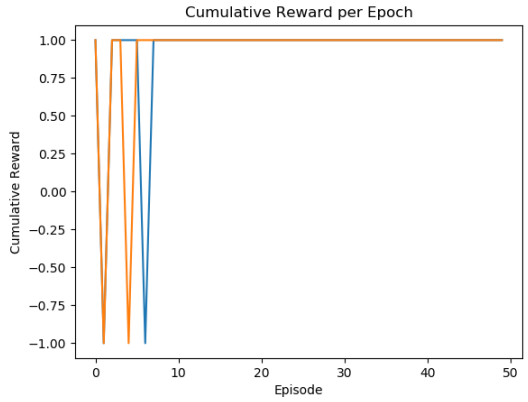


Figure 1: Cumulative Reward per Training Episode

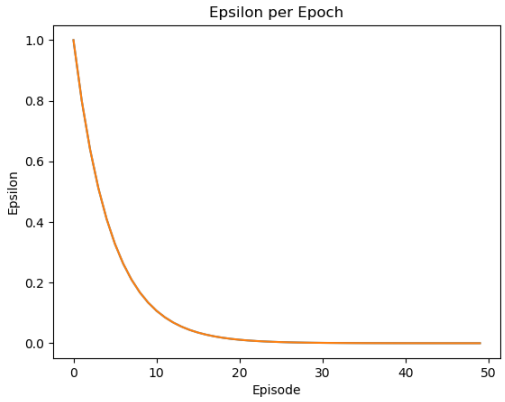


Figure 2: Epsilon per Training Episode

Figure 3: Learned Agent Movements