Pranshav Thakkar

pthakkar7

CS 4731

**Homework 8 Report**

1. The rewards system is set up so that being in radiation costs you a one-time cost of 20 points, while saving a person gives 20 points per turn. So, it is beneficial for the agent to try and save all the humans. In the default map, there is a clear path to the first human that needs saving, but the second human is blocked by radiation and there is no other way to get there. There shortest path to thee first human is going through one block of radiation, and the agent, after enough iterations, learns that it is worth going through the block of radiation to save the human, as the cost of doing that will be offset by saving the human and it will actually gain more positive reward in the future by saving the human. Likewise, the agent learns that it is worthwhile to cross the radiation on the right side of the map to save the second human, as the one-time costs of being in the radiation are not significant compared to the positive rewards gained by saving the humans.

For the robot to avoid the radiation, there would have to be a high enough negative cost for it to not be worth passing through, or the cost for saving the humans must not be high enough to offset the cost from the radiation, or there must be a path for the robot to reach the second human without there being radiation in the way.

1. For this problem, let us consider the situation that there are no humans to be saved and there is just the interaction between the agent and the enemy. In this case, the agent would attempt to move to the goal and get the 10 points for staying on the goal, while avoiding the enemy. So, lets say it gets to the goal and it spends about half the time on the goal and half the time avoiding the enemy. Let’s say it takes 10 turns for this cycle to repeat. Then, 10 – enemyDead would have to be greater than (10\*5 – 5\*5)/10. So, enemyDead would have to be less than 5 for it to be worthwhile for the agent to kill it while passing, because any other higher value would put the agent in an overall loss. Obviously, enemyDead has to have a negative cost so it would be -5, because a positive value for enemyDead would make it beneficial for the agent to kill it.

The smallest value for enemyDead that would cause the agent to chase and kill the enemy would be any positive float value. It can even be 0.001 or something similar. This is because killing the robot would add a positive reward every turn after that, and that can only add to the overall reward. It also helps because then the agent does not have to waste turn trying to avoid the robot and incurring the -1 penalty for moving if it is not on the goal.

1. Switching to mode 2 causes the enemy to move randomly, and so training for 500 episodes (the standard) or fewer is not enough for the agent to learn the enemy’s behavior as compared to the influence map, so the agent just tries to avoid the enemy as best as possible. However, with higher training episodes, like 1000 or more, we start to see that the agent understands that the enemy is moving randomly and is likely not to stay in one spot for more than one turn. So, overall it tries to avoid the radius of the enemy, but in addition it is able to make predictions based on the fact that the enemy will move. So, if the enemy is blocking the goal on one turn, the agent will still move to the goal on the next turn because it will predict that the enemy will move away from it and this can help accumulate more overall reward over time.