

HW8 Report

1. In the default map, the agent learns that the point loss of entering the radiation is worth rescuing the people. Each person saved gives a large bonus per turn, and the agent will learn this as it makes moves early on in training and happens to save a person. The agent will learn that risking entering radiation and losing points to reach a person is better than avoiding that path. Since killing the enemy causes loss of points, the agent will learn this and decide to avoid the enemy and keep it alive, entering the radiation to save the people. To reach the second person, the agent must enter radiation, since all paths to that person are blocked by a wall or radiation square. After many iterations, the agent learns to move through the radiation to reach the person because the reward per turn of saving them is worth it.

Multiple changes to the map would cause the agent to learn to avoid radiation. Firstly, moving the second person to the left side of the map would result in the agent not needing to move through radiation to rescue people. Moving the second person or removing one of the rightmost radiation squares (or the wall) would allow the agent to save both people without touching radiation, and since there is a negative reward for being in radiation, the agent would learn to avoid it since it is not a requirement to maximize its points.

Additionally, simply making the reward of being on a radiation square very negative will cause the agent to avoid radiation to keep its score high. If radiation rewards -2000 or something similar, the agent will learn to avoid it.

2. The smallest reward that will make the agent OK with killing the enemy is 0 for enemyDead. If the reward is negative, the agent will learn not to do so in order to maximize points. If there is no loss when the agent kills the enemy, then the agent will simply kill the enemy when they meet. However, the agent will not seek out the enemy since there is no positive reward to killing it, just that it is OK to kill the enemy when they meet.

To make the agent seek out the enemy, a reward of 21 is required for enemyDead. This is greater than the reward for saving a person, meaning that the agent will seek out the enemy first, and the people only after killing the enemy. Once the enemy is dead and the agent is gaining 21 points per turn, it will move on to the next highest priority task (which will be saving the people).

3. In mode 1, the agent can fairly quickly learn the enemy's pattern of movement and operate accordingly, since it is simply following the influence map at every location. In mode 2, the player's control of the enemy is more random than this influence map, so the agent has a harder time predicting its opponent's moves and will make suboptimal moves. This will cause the total reward for the player agent to increase as compared to mode 1. Increasing the number of training episodes will create a higher maximum reward amount, because the agent can reach more states. As the agent does so, it will become closer and closer to the optimal policy. Seeing all the possible states will allow the agent to find the optimum policy, so increasing the number of training episodes will move the agent closer to the correct moves while reducing this number will have the opposite effect.