CS 4731 Game AI HW8 report
An Jihai (jan61)

Question 1:
Why doesn't the bot avoid the radiation in the default map? What would have to be different for the bot to avoid as much of it as possible?

The reinforcement learning algorithm uses the reward as the criteria of learning. The reason why the bot avoids the radiation in the default map is that it thinks it will earn more if it goes through the radiation. Let's check our current rewarding system. +20 per human rescued, +10 for being at the base station, -20 for passing the radiation, and -1 for each step. Let's see the human with id = 7 as an example, because he is trapped in an area surrounded by walls and radiations. In the default map, the agent will try to save human 6 at first. Before returning to the base, this agent will also try to rescue human 7. When the agent wants to rescue this human 7, it will need to pass a radiation block, then pick up the human, finally cross the radiation area again and then return to the base. Mathematically, the action of crossing the radiation area and saving the human 7 with n steps will return a total reward (-20+20-n). After picking the human 7 up, the agent will cross the radiation area again and return to the base. But since we have a discount value in our formula, the agent won't notice this lost at first(actually, the agent will only notice a small portion of this lost, for example, $-20*0.8^4$). In one word, the agent with the default rewarding system thinks crossing the radiation area will earn itself more than avoiding the radiation area, therefore it won't avoid the radiation. To make the bot avoid radiations, simply increase the penalty of being in the radiation. For instance, I set the penalty of being in the radiation as -800. With this setting, the bot will never try to cross the radiation area since it knows it will lose 800 points(basically all points it has.) for being on that block. Under this rewarding system, the bot will go down first and then go right to avoid the radiation area, then rescue human 6, and finally return to the base. Human 7 will never be saved since he is in an 'unreachable' area(surrounded by walls and radiations).

Question 2:
Under the default reward, the bot runs away from the enemy. What is the largest value for enemyDead that would make it so that the bot is willing to kill the enemy if they cross paths. Explain why. What is the smallest value for enemyDead that would induce the bot to seek out the enemy and kill it. Explain why.

The original reward for killing the enemy is -10 and under this penalty, the agent won't kill the enemy since it will lose too many points by killing it. I increase the reward by one on each try and I found when the reward is -2, the bot will kill the enemy intentionally. However when the reward is -3, the bot will still ignore the enemy. So I may conclude that the smallest value for enemyDead that would make it so that the bot is willing to kill the enemy if they cross is -2. Under this value, the agent will stay in the base, and kill the enemy if the enemy is also in the base. To see the correctness of this value, we can do some simple calculation. With the default setting, the agent will earn 20+10 = 30 points at each turn, without killing the enemy. After change the reward of killing an enemy to -2, the agent will earn 30-2 = 28 points at each turn. We can see the difference is really small, after taking the gamma value into account. Therefore the value -2 here is reasonable.

There are two ways to earn points in this game: one is being in the base area, the other is rescue a human. The first option will earn +10 points, and the later option will earn +20 option. However, after reading the map, we found that the reward of saving a human will be nullified by the penalty of crossing the radiation area. To let the agent seeking for a kill, I set the reward a little higher than the reward of being in the base. Therefore the smallest value is 10+1 = 11.

Question 3:
What effect does switching enemyMode from 1 (follow the influence map) to 2 during training have on the behavior of the bot, if any? How does more or less training episodes help or hurt? Hint: experiment with play = 2.

When the enemyMode is set as 1, the enemy is moving as the influence map indicated. Therefore the agent is trained specifically according to this fixed behavior of the enemy. The model after training will work better against an enemy moving in this path. When the enemyMode is set as 2, the enemy is moving randomly and unpredictable. This situation is more like a human player in controlling the enemy. So the trained model under this enemyMode will be better in handling enemy's random moving path. Therefore after changing the enemyMode from 1 to 2, the trained model's agent will be stronger against randomly moving enemy(for example, a human player). When the enemyMode is 1, the enemy will follow a certain path. Under this situation, the Q-value for some states will remain unchanged after a certain number of rounds of training. In other word, the v_table will converge eventually. Thus, more training episodes won't help if the v_table has converged. When the enemyMode is 2, since the enemy is moving randomly, the table has no way to converge. So, the more training it has, the more situation it will learn. Thus more trainings will always help.