After considering Dr. Doshi’s advice, I have some new general ideas now. The environment is still set to be a mini-game of StarCraft II, where two cooperative units controlled by the agent are collecting mineral shards (static neutral units) randomly distributed in the map while scouting. The value of a policy is counted by the number of shards collected by the two units within certain game time. I am going to utilize a model-based reinforcement learning method where utilities are learned by adaptive dynamic programming. The challenge here is that how to adjust the method to better adapt to continuous setting in the RTS game. Computational complexities of time and space are also need to be taken into consideration.