Anubhav Sharma, 2018114007
Rishabh Daga, 2018101015

Report

Team 23

Problem Statement

Here we again have to implement **Value Iteration Algorithm** to get the final optimum policy, for which further we have to find the maximum value of rx such that the given constraints hold true.

$$\max(\mathbf{rx}) \mid \mathbf{Ax} = \alpha,\ \mathbf{x} \geq 0,$$

So for further doing the maximization problem we would need to find the matrix A.

A Matrix

The A matrix has a dimensions of n*m where n is the number of states and m is the number of all possible actions from every state. In the given problem there were in total 100 possible actions for all 60 states.

Now for each state that we traverse, a functionality is used where we try every action for that state in order and if find that a particular action can be executed from a state then we add a new column to A matrix where the value corresponding to outgoing sate is given a +1 and the ingoing states are given a value corresponding to probability of reaching that state.

Actions

The action described over here follows the order in which they are firstly possible.

- ## NOOP

Only states with Enemy Health 0 can undergo NOOP action. They can neither undergo any other action **nor can any other state** undergo this action.This action if possible has to be executed first.

- ## Recharge

Recharge action is performed only if the stamina is not equal to MaxStamina and after taking that action probability values corresponding to the new states are added to the A matrix.

- ## Dodge

Dodge action is performed only if the stamina is non zero and after taking that action probability values corresponding to the new states are added to the A matrix.

- ## Shoot

Shoot action is performed only if the stamina as well as Number of Arrows are non zero and after taking that action probability values corresponding to the new states are added to the A matrix.

So for all the states, feasible actions are added to the A matrix.

Final Policy and its Selection

After calculating A matrix, we calculated the reward matrix and the alpha matrix. As Our team number is 33 we used -20 Penalty.

The Reward matrix consists of 0 for NOOP action and -20 otherwise.

The alpha matrix denotes the probability of initial state,which is 1 for the last state (4, 3, 2) and 0 otherwise.

Then we used Linear programming for finding Maximum (Reward * x)

Under Constraints,

Ax == alpha , x>=0