

MDL Assignment 2 Part 2

Team 24

Pooja Desur 2019101112

Manasvi Vaidyula 2019101012


Part 2 : Value Iteration Algorithm

VI Algorithm

The Value Iteration algorithm runs through many iterations, updating the utility value of every single state in the Markov Decision Process until a set convergence (Bellman error) is reached.

In every iteration, each state's utility value is updated based on the future actions that can be taken from that state and current utility of future states that can be reached upon taking those actions. For each state, the maximum value of the sum weighted probability of possible states that can be reached across all actions that can be taken from that state is the new utility of that particular state.

The current utility of each state being updated is known as the Bellman Update -


$$U_t(s_i) = \max_A \left(\sum_{j=0}^k P(s_i, a, s_j) * (R(s_i, a, s_j) + \gamma * U_t - 1(s_j)) \right)$$

In this algorithm, there are a set of step costs : (-20, -60, 30). These are taken into account based on which future state a state can reach by taking a particular action.

TASK 1

Results + Analysis

The algorithm finally output a policy which determined the best action to be taken from each state that would maximise Indiana Jones's final reward.

All the states where the Monster's (MM) health is zero signal termination of the game, hence all these states' best action is NONE, no matter the position Indiana Jones (IJ) is at.

When IJ is at C, and the monster is Dormant, the best action is usually always RIGHT, which would take him to the East position where he can HIT and SHOOT with a better probability of hurting MM, no matter the number of arrows and materials IJ currently has. When MM is at a Ready state and IJ is at C, the best action is UP, SHOOT, or RIGHT depending on the number of arrows IJ currently has, and the health of MM. When MM has 25 health and has at least one arrow, the best action is SHOOT, as with one successful shot IJ can win the game, otherwise, the best action is UP, leading IJ to N position. At N, MM cannot attack IJ, so this state is preferable to IJ compared to position C and E where he can be attacked. If IJ has no arrows, the best action would be UP for the same reason. The amount of material IJ has doesn't play any role in why an action is the best action for this position.

When IJ is at N, if MM is in Dormant state, when IJ has no material to CRAFT any arrows, the best action is DOWN, so he has a chance at attacking MM. If MM is Ready, it would not be wise to go to a position where he can attack IJ, so best action is STAY. When IJ does have materials to CRAFT arrows, when he has 0 arrows, the best action is to CRAFT. This would be so he has ammunition in order to SHOOT MM at C and E position later on. When MM is Ready, the best action is to STAY or CRAFT, in order to avoid going to a position where IJ can be attacked. The more arrows IJ has, and MM is in Dormant state, the best action is more likely to be DOWN than CRAFT, as he already has a larger set of arrows. IJ having 1,2, or 3 materials to craft arrows does not influence the best action.

At S, when MM is Dormant, the best action is always UP. By taking this action, IJ gets to C position, where he can attack MM. However when MM is Ready, the best action is influenced upon the number of materials and arrows IJ currently possesses. It is better to STAY at S, than to travel to C where MM

could attack. However when IJ has 0 arrows, at certain states, it is the best action to move UP, so that they can end up at N position eventually to CRAFT arrows and be able SHOOT MM later on..

If IJ has zero materials and arrows, the best action is to GATHER in order to CRAFT more arrows later.

At W, when MM is Dormant, the best action is usually RIGHT. This is so he can get to C or E where he can attack MM by SHOOT or HIT with a better probability of success. At W, IJ can only SHOOT with very low probability of success. However this is dependent on the number of arrows IJ has currently. When he has 3 arrows, the best action is to SHOOT. This is because even if he misses, he has more arrows to attack with. In all other cases the best action is always RIGHT. When MM is Ready, the best actions vary between STAY, SHOOT, or RIGHT depending on the number of arrows and health of MM. When he has no arrows, the best action is RIGHT, probably so he can CRAFT more with material at N.

At E, when MM is either in Dormant or Ready state, the best action is always that of attack - either SHOOT or HIT. Since at this position all actions of attack are most likely to be successful than any others, no other actions are preferred here. When IJ has 0 arrows, he always HITS in order to attack the monster. SHOOT has a higher probability of success (0.9), but deals less damage (-25) whereas HIT has a lower probability of success (0.8) but deals more damage (-50). Therefore when MM's health is 100, IJ's best action would be to HIT, but in all other cases his best action is to SHOOT. The only case where he would choose to SHOOT rather than HIT at health 100 is when he has 3 arrows and MM is Dormant. Amount of material he has doesn't influence his decision for best action here at all.

Rate of Convergence

The algorithm converged after 126 total iterations. The Bellman error was 0.001 (the convergence point where the algorithm only stops when the difference between all states updated utility and previous utility is lesser than it). Gamma, which used to decided whether more weight should be put towards short-term

of long-term gains, was set to 0.999. This value is very close to one, and thus it means this particular policy optimizes the gains over infinite time.

Simulating the Game


Procedure :

The player follows the policy derived from Task 1. At every state based on the best action (from the policy) it randomly picks from the possible neighboring states and the process continues till a terminal state (Monster health = 0) arrives.

1. Start State : (W, 0, 0, D, 100)

 **Current State :** W 0 0 D 100 **Action :** RIGHT **Next State :** C 0 0 R 100


The policy says RIGHT is the best action for IJ since MM is in a Dormant state and going to C would help IJ attack MM with a better success probability using SHOOT or HIT actions. He can also use HIT action which he can't use in W, which deals more damage to MM terminating the game soon.

 **Current State :** C 0 0 R 100 **Action :** RIGHT **Next State :** E 0 0 R 100

The policy says RIGHT is the best action for IJ and going to E would help IJ attack MM with a much better success probability in both SHOOT and HIT actions.

 **Current State :** E 0 0 R 100 **Action :** HIT **Next State :** E 0 0 R 50

The policy says HIT is the best action for IJ at this point, as it deals more damage than SHOOT, though it has a lower success probability, and will kill MM.

 **Current State :** E 0 0 R 50 **Action :** HIT **Next State :** E 0 0 R 0

The policy says HIT is the best action for IJ at this point, as it deals more damage than SHOOT, though it has a lower success probability, and will kill MM immediately.

 **Final State :** E 0 0 R 0

2. Start State : (C, 2, 0, R, 100)

 **Current State :** C 2 0 R 100 **Action :** UP **Next State :** C 2 0 D 100

The policy says the best action is UP but because we chose states randomly. Here the transition was the MM attacked IJ, so the action was unsuccessful and R changed to D.

 **Current State :** C 2 0 D 100 **Action :** RIGHT **Next State :** E 2 0 R 100

The policy says Right (Move to E) is the best Action for IJ at this point, because at E he can hit and shoot him with a higher probability.

 **Current State :** E 2 0 R 100 **Action :** HIT **Next State :** E 2 0 D 100

The best action is HIT, for the same reasons listed above. Here the MM attacked, rendering the action unsuccessful, which was a randomly chosen state.



Current State : E 2 0 D 100
50

Action : HIT

Next State : E 2 0 R

The policy says HIT is the best action for IJ at this point, as it deals more damage than SHOOT, though it has a lower success probability, and will kill MM immediately. Hit success and MM changing to D state were randomly chosen.



Current State : E 2 0 R 50
R 50

Action : HIT

Next State : E 2 0

The policy says HIT is the best action for IJ at this point, as it deals more damage than SHOOT, though it has a lower success probability, and will kill MM immediately. The randomly chosen state was that where HIT wa



Current State : E 2 0 R 50
R 50

Action : HIT

Next State : E 2 0

The policy says HIT is the best action for IJ at this point, as it deals more damage than SHOOT, though it has lower success probability, and will kill MM immediately. The randomly chosen state was that where HIT was unsuccessful.



Current State : E 2 0 R 50
D 75

Action : HIT

Next State : E 2 0

The policy says HIT is the best action for IJ at this point, as it deals more damage than SHOOT, though it has lower success probability, and will kill MM immediately. The randomly chosen state was that where MM attacked.



Current State : E 2 0 D 75
D 25

Action : HIT

Next State : E 2 0

The policy says HIT is the best action for IJ at this point, as it deals more damage than SHOOT, though it has lower success probability, and will kill MM immediately. The randomly chosen state was that where H.



Current State : E 2 0 D 25
D 0

Action : HIT

Next State : E 2 0



Final State : E 2 0 D 0

TASK 2

Case 1

Indiana now on the LEFT action at East Square will go to the West Square.

Results + Analysis

The best actions in the policy generated are the same as the policy generated in Task 1.

Rate of Convergence

The algorithm converged after 127 iterations, which is very similar to the original algorithm in Task 1. Since the step cost nor gamma value was changed, changing the way a particular action worked for a particular position did not affect the convergence rate by too much.

Case 2

The step cost of the STAY action is now zero.

Results + Analysis

The best actions for all positions changed dramatically from the policy in Task 1.

When IJ is in C position, in most states, his best action is to move LEFT. This is because on moving to W position, he can attack MM, without MM being able to attack him. STAY at W will not cost him anything (step cost is zero for that action), so he can STAY at W and SHOOT from that position. If IJ has at least one arrow, and MM health is 25 the best action is not LEFT but rather SHOOT as he can kill MM if that shot is a success and end game getting +50 reward. If MM health is higher, it would be harder to get two successful shots in a row, so the best action would be to go LEFT and SHOOT from W instead.

At N, IJ's best action is DOWN when he has no material, no arrows and the monster is Dormant, similar to Task 1. However when MM is Ready, he will always choose to STAY or CRAFT and remain in N position where MM cannot attack him. In Task 1, sometimes the best action was to move DOWN over STAY when he had no material. Here, he will never go DOWN. When MM is dormant, he will either go DOWN, STAY, or CRAFT. DOWN will help him reach states where he is in W position.

At S, just like in N position, when MM is ready, IJ will always choose to STAY as it has the least step cost, and he won't move to a position where it is possible for MM to attack him, getting him a negative reward. When MM is Dormant, he either STAY or UP, so he can reach state where he is in W position.

At W, the best action is to always STAY. He cannot be attacked by MM here ever. The step cost of SHOOT would be worse than that of STAY, so the policy dictates to always STAY.

At E, similar to in C position, a lot of the state's best action is to move LEFT, so he can eventually reach W position, where the policy always dictates the action where step cost is very high in comparison. If MM health is almost depleted, and IJ has arrows, the best action is to SHOOT so the game can terminate. The

policy here is very different from Task 1, where the policy at all states with E was to attack with either SHOOT or HIT.

Rate of Convergence

The step cost increased from -20 to 0 for many of the actions. The algorithm ran in 64 iterations, much quicker as it was easier to reach a higher reward state with this higher step cost.

Case 3

Change the value of gamma to 0.25.

Results + Analysis

With this lower gamma value, IJ was focusing on short term gains, and not worried about the long term.

At C, the best action would be to HIT or SHOOT much more often than in Task 1 policy. In fact Task 1 policy never even had HIT in any states at C, whereas here, the best action is to HIT even in some states where IJ has arrows (HIT deals more damage).

At N, when IJ had no material, and MM was R, depending on the number of arrows IJ had the best action would be to DOWN or STAY. By going DOWN, he could attack MM. There were more CRAFT best actions in this policy than in Task 1, even when IJ had the maximum number of arrows. This is good avoidance technique when MM is Ready, and MM cannot attack IJ when he is in N position.

At S, there was a lot more GATHER in the policy than in Task 1. This is another good avoidance technique as MM cannot attack him at S position. When MM is at R, it is preferable to GATHER even when IJ has max materials. When MM is D, the policy included UP, so IJ could get to a position where he could attack.

At W, there was much more SHOOT than in Task 1. Since he was only looking at short term goals, IJ was trying to end the game sooner by attacking from W, though this had a lower success probability than in C or E. If he had even one arrow he would SHOOT, else he would STAY when the MM was Ready (to avoid attack as he can't attack at W position) or RIGHT so he could CRAFT more arrows or HIT MM.

At E, the policy had much more HIT than SHOOT, which was different from Task 1 policy. HIT dealt more damage, though it had a lower probability of success. But since this algorithm was tailored toward short term goals, a quicker termination of the game was desirable. IJ only chose to SHOOT when MJ had 25 Health, and at least one arrow, as it would have a higher probability of success than HIT, and both means of attack would lead to termination of the game.

Rate of Convergence

By decreasing gamma by a very significant amount compared to the original algorithm in Task 1, it focused more on short term optimization. The algorithm only ran for 10 iterations and converged more than 12 times quicker.

The policy says HIT is the best action for IJ at this point, as it deals more damage than SHOOT, though it has lower success probability, and will kill MM immediately. The randomly chosen state was that where HIT was unsuccessful.