MDL Report

Team 79

Snehal Kumar - 2019101003

Ashish Gupta - 2019101061

Assignment 2

Value Iteration

It is a method of computing an optimal MDP policy and its value. Value iteration starts at the "end" and then works backward, refining an estimate of either Q* or V*.

The Bellman equation is the basis of the value iteration algorithm for solving MDPs.
Let Ut(I) be the utility value for state s at the t'th iteration. The iteration step, called a Bellman update, looks like this:

$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s') \right)$$

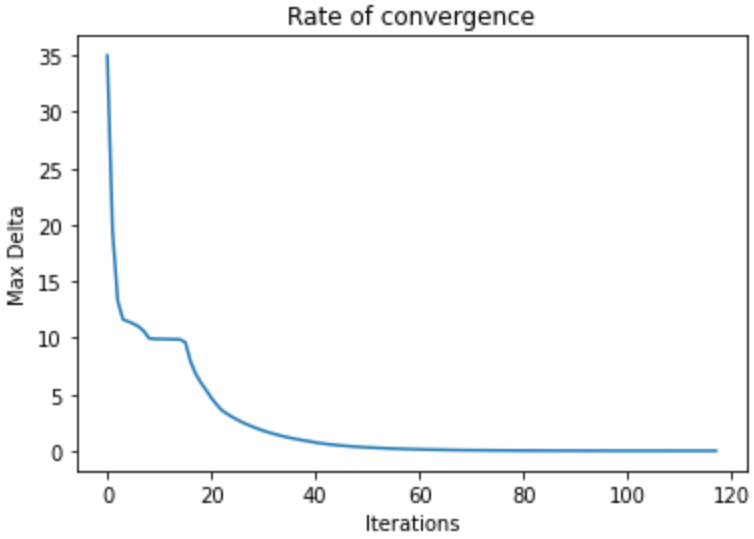
Task 1: Value Iteration

Given below are few observtions we derived from the trrace.txt file:

- It took 117 iterations for the algorithm to converge

- Indiana Jones is not a rash decision-maker. Due to the high value of gamma, it puts more weight on the long-term goals. This can be inferred from the fact that IJ prefers to move away from/dodge MM whenever it is in its ready state
- Being a risk-averse person IJ does not prefers to shoot whenever the possibility of hitting MM is low or has a low amount of arrows. The opportunistic behaviour is highlighted when he prefers to move away from MM whenever he has a chance and MM is in a **ready** state and craft arrows or gather materials depending on the number of utilities left
- IJ prefers to use his blade from the Eastern box due to the high amount of damage possible, especially when MM's health fully recharges.
- Whenever in the central box and MM is in the dormant state,IJ prefers to move right and then shoot mm with his arrows/blade so as to get a higher probability of hitting him.
- Due to the Step cost of -10, IJ Uses his **STAY** Option only in South, North, and in Western Square only to escape from the ready state of MM.
- Although IJ doesn't explicitly go to the South to Gather materials, it is found that whenever IJ tries to move away, it goes to South in case of less materials to avoid MM as well as Gather

Rate of convergence:



Simulation

1. **(W, 0, 0, D, 100)**

Simulation started from ('W', 0, 0, 'D', 100)

STATE	ACTION	RESULT	MM_ACTION	MM_ACTION RESULT
('W', 0, 0, 'D', 100)	RIGHT	Success	Not attack	DORMANT
('C', 0, 0, 'D', 100)	RIGHT	Failed	Not attack	DORMANT
('E', 0, 0, 'D', 100)	HIT	Failed	Not attack	DORMANT
('E', 0, 0, 'D', 100)	HIT	Success	Not attack	DORMANT
('E', 0, 0, 'D', 50)	HIT	Failed	Not attack	DORMANT
('E', 0, 0, 'D', 50)	HIT	Failed	Not attack	DORMANT
('E', 0, 0, 'D', 50)	HIT	Failed	Not attack	DORMANT
('E', 0, 0, 'D', 50)	HIT	Failed	Not attack	DORMANT
('E', 0, 0, 'D', 50)	HIT	Success	Not attack	DORMANT

2. (C, 2, 0, R, 100)

Simulation started from ('C', 2, 0, 'R', 100)

STATE	ACTION	RESULT	MM_ACTION	MM_ACTION RESULT
('C', 2, 0, 'R', 100)	UP	Success	Not attack	READY
('N', 2, 0, 'R', 100)	CRAFT	Success	Attack	FAILED
('N', 1, 2, 'D', 100)	DOWN	Success	Not attack	DORMANT
('C', 1, 2, 'D', 100)	RIGHT	Success	Not Attack	DORMANT
('E', 1, 2, 'D', 100)	HIT	Failed	Not Attack	DORMANT
('E', 1, 2, 'D', 100)	HIT	Success	Not Attack	DORMANT
('E', 1, 2, 'D', 50)	SHOOT	Success	Not Attack	DORMANT
('E', 1, 1, 'R', 25)	SHOOT	Failed	Attack	SUCCESS
('E', 1, 0, 'D', 50)	HIT	Failed	Not Attack	DORMANT
('E', 1, 0, 'D', 50)	HIT	Failed	Not Attack	DORMANT
('E', 1, 0, 'D', 50)	HIT	Failed	Not Attack	DORMANT
('E', 1, 0, 'D', 50)	HIT	Success	Not Attack	DORMANT

> In both cases, we see that on running a simulation with random probability, it ends up in the end state.

Task 2

Case 1:

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

The number of iterations slightly increased to 119.

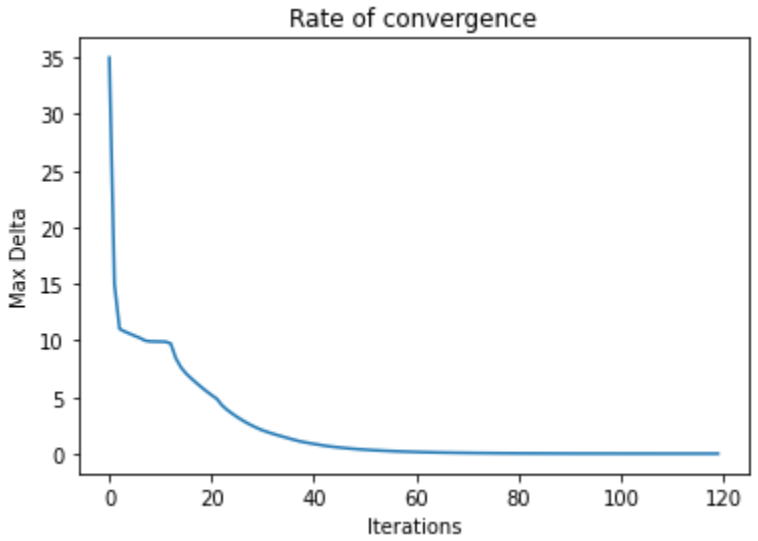
Change in Policy:

IJ can now escape from the attacks of MM whenever in a ready state and does so when he has a high number of arrows in the fear of losing them.

Analysis:

The policy remains pretty much the same as of normal one except for the four changes when IJ chooses to move LEFT to evade from MM's attack.

Rate of convergence:



Case 2:

> The step cost of the STAY action is now zero.

The number of iterations now dropped down from 117 to 56.This may be due to the following reasons

Change in Policy:

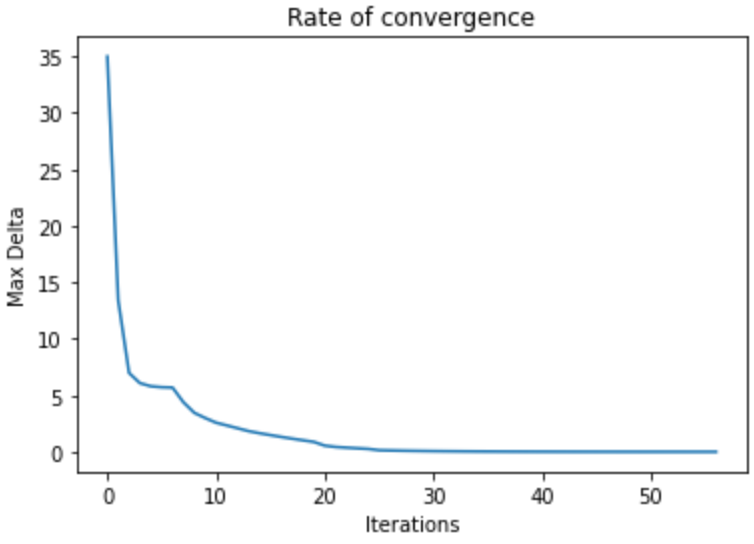
The frequency of choosing **STAY** increases since the step cost of STAY is zero. The agent becomes risk averse and tries to avoid MM by moving towards West.

Analysis:

Due to Decrease of step cost of Stay option the frequency by which IJ chooses to stay increases

IJ now has become extremely risk averse.This behaviour can be prompted by the fact that it tries to run away (moves LEFT) from a **READY MM** unless there is good chance it will **kill** MM in that step. Whenever it gets the chance, it goes to the West square by choosing **LEFT** action most of the times when it is in Center square to avoid MM. Other times, it tries to **STAY** in North and South as the step cost is zero.

Rate of convergence:



Case 3:

> Gamma = 0.25

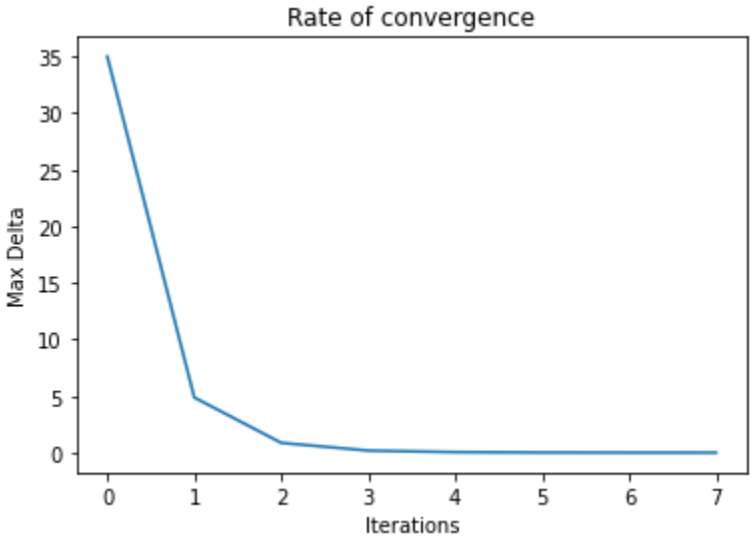
Change in Policy:

The agent now has become risk-neutral and misses the long-term effects of its actions. The frequency of choosing to **attack** MM increases to get to its goal as fast as possible and get the reward.

Analysis:

- The number of iterations went drastically down to 7.
- This is expected as, with decrease in the value of gamma, short term gains are prioritized and IJ is unable to lookahead too many states and thus, aims to kill MM as soon as possible. IJ has now become risk neutral
- This behaviour can be shown as he aims to attack MM whenever he gets a chance.
- However still, IJ chooses to stay away from a Ready MM whenever he has a chance.
- Whenever in the eastern box IJ prefers to use his blade than his arrows in order to get a high damage.
- He also chooses to use his arrows if MM's health is 25 since they have a higher chances of hitting successfully.
- On running simulations, it is found that, due to IJ's change in behaviour, MM ends up surviving many times as IJ moves to stay away from MM.

Rate of convergence:


Rate of convergence
Max Delta
Iterations
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6
7