

# Policy Iteration and Value Iteration\*

An Khanh Vo

19520007@gm.uit.edu.vn

May 2021

## Abstract

In this report, we try to find the solution for FrozenLake-v0, FrozenLake8x8-v0 and Taxi-v3 environment using Policy Iteration and Value Iteration. We have implemented these algorithms and compare their performance. This experiment we run entirely on Jupyter Notebook.

## 1 Result Analysis

In FrozenLake-v0, Policy Iteration proved to have performance than Value Iteration with a result of 733/1000 successful episodes. Value Iteration only won 720/1000 episodes.

In FrozenLake8x8-v0, with state space expended from 16 to 64 compared than FrozenLake-v0, Policy Iteration has clearly won over Value Iteration than last time with 756/1000 successful compared to 738/1000.

Taxi-v3 is a slightly different environment from FrozenLake. The probability when in state  $s$ , taking action  $a$  and arriving at state  $s'$  is 100%. Another different is that it encourages the agent to reach the goal as soon as possible by penalizing the agent 1 point per time step elapsed. The results of both algorithm are very good when both succeeded 1000/1000 episodes.

## 2 Conclusion

Both Policy Iteration and Value Iteration can work efficiently when the agent is assumed to have prior knowledge about the effects of its on the environment. They will work even more effectively in deterministic environments, such as Taxi-v3.

Policy Iteration has less computational cost than Value Iteration because it has much fewer iterations, but each iteration of Policy Iteration is much more expensive than Value Iteration.

---

\*This is my report for CS106 - Artificial Intelligence (Spring 2021) at University of Information Technology - Vietnam National University HCMC