Value Iteration and Policy Iteration

Truong Minh Chau

June 2021

Contents

1 Value Iteration

In Value Iteration, we start with a random value function, then find a new value function in an iteration until reaching the optimal value function and derive the optimal policy from it. Finding optimal value function can also be seen as a combination of policy improvement and truncated policy evaluation.

2 Policy Iteration

In Policy Iteration, we start with a random policy π, then find the value function of that policy through the Policy Evaluation function and continue to improve to have a new policy based on the previous value function by using the Policy Improvement function. The two steps are repeated iteratively until policy converges.

3 Run on OpenAI Gym environments

We test how long these two algorithms will converge at as well as the number of successful times they can perform for 3 different environments below.

3.1 FrozenLake-v0

Algorithm	Converged at	Successful time
Value Iteration	79	720
Policy Iteration	5	709

3.2 FrozenLake8x8-v0

Algorithm	Converged at	Successful time
Value Iteration	117	737
Policy Iteration	9	730

3.3 Taxi-v3

Algorithm	Converged at	Successful time
Value Iteration	116	1000
Policy Iteration	16	1000

4 Summary

In conclusion, the Policy Iteration works on principle of " Policy Evaluation $\longrightarrow$ Policy Improvement" whereas Value Iteration bucks the principle of "Optimal Value Function $\longrightarrow$ Optimal Policy". These data sources clearly communicate

that the Policy Iteration runs less iterations to converge while the policy extracted by both algorithms has a roughly similar data. For each iteration the Value Iteration has a better time complexity.