CS 312: Artificial Intelligence Laboratory

Task 8: Reinforcement Learning

Goal: Formulate and solve an MDP problem using policy and value iteration.

Note: This assignment is to be done individually.

(1) Come up with a sequential decision-making problem and formulate the problem as either a discounted reward or total reward MDP problem.
(2) Develop code for solving the MDP problem using policy and value iteration.
(3) Write a report clearly describing the MDP considered and your observations on running the policy and value iteration algorithms on the formulated MDP.
(4) Further, one should also suggest ways to check whether the algorithm yields optimal policy for the setting considered.

Note:

(1) The MDPs considered should not be the same across students. There should be a significant difference in the MDP considered.
(2) Do not take MDPs that have a large number of states/actions. It is not possible to solve them then.
(3) In order to meet Instruction 1 above, each student should decide the MDP they want to consider as early as possible and fill in this sheet.
(4) If someone has already taken that MDP problem you are supposed to decide on a new MDP problem.
(5) Hence, students quickly deciding on their MDP problem will have less chance of changing the MDP problem again.

You should take the consent of the TAs by filling this sheet to finalize on the MDP problem at the earliest.

Evaluation Criteria:
MDP formulation: 10
Correctness: 20 (10 + 10 for PI and VI)
Report: 15
Code Quality: 5

Deadline: 11:59 PM 6 April 2020

Late Submission Policy: 5% of marks will be deducted per day late.
For Reference :

Reinforcement Learning:
http://www.cse.iitm.ac.in/~ravi/courses/Reinforcement%20Learning.html
See lectures 15 -25.

Report Format :

1. [1 mark] MDP Description: Clearly describe (S, A, P, R, N)
2. [5 marks] State-transition Graph for the MDP
3. [1 mark] Optimal Policy: Suggest ways to check whether the algorithm yields optimal policy for the setting considered.
4. [5 marks] Experimental Results: Vary the gamma parameter, show the policy found in each case by both algorithms
5. [2 marks] Comparison of Policy Iteration and Value Iteration
6. [1 marks] Conclusions