FILA: Assignment-2 Report

Ashish Aggarwal
Roll no. : 203050015

Monday 19th October, 2020

1 Task 1 - MDP Planning Algorithms

1.1 Value Iteration

Started with initial value function V_0^* for all states as zero

1.1.1 Assumption

Tolerance for the vi algorithm is 1e-8.

1.1.2 Observations

This is working best over wide range of MDP formulations.

1.2 Howards Policy Iteration

Started with initial policy π_0 taking first action for all states i.e. a=0.

1.2.1 Implementation and Design

Policy Evaluation: After some waving off with the Bellman equation I got the following equation and used this to do policy evaluation. I have used Numpy library to solve linear equations:

$$(I - \gamma\, T(s_0, \pi(s_0), s_0))V(s_0, \pi(s_0), s_0) - \sum_{s \in S, s \notin S_0} \gamma\, T(s, \pi(s), s)V(s, \pi(s), s) = \sum_{s \in S} T(s, \pi(s), s)R(s, \pi(s), s)$$

Policy Improvement: I have used action which gives highest action value for a given state.

1.3 Linear Programming

1.3.1 Implementation

Added n variable to set of linear equations and added n*k constraints to LP problem for optimization.I have used pulp library for this purpose. I chose minimization of element wise sum of value function as my objective function.

2 Task 2 - Solving a maze using MDPs

2.1 Implementation

I have added no of states equal to no of element in grid file which are not equal to 1 and 4 actions each for left, right, top, down movement.All transitions have probability of 1.

- Case1: empty to wall movement: added reward value of -100000. I thought it will converge fast if we'll give large negative rewards.Here state is not changed.
- Case2: empty to empty movement: added reward value of -2 negative because we have to find shortest path.In case of positive value It can keep oscillating around a empty cell still gain rewards.we can choose any non positive value.State will be switched after transition.
- Case3: empty to end: added reward value of 1 only positive value will work as a stopping criteria to maximize reward.State will be switched after transition.

Apart from this all other transition probabilities are zero. I have maintained a state to position in grid dictionary to decode/encode.