CS747 Assignment-2 170050067

Value Iteration:

Algorithm:
Calculating Vt for each state, till Vt+1-Vt <precision
The precision used was 0.0000001
Obtained the policy, from relationship between V*,Q*,Pi*

Howards Policy iteration:

From transitions, made a arbitrary policy(took one of action that has a transtion)
After computing Q(s,a) for each action, swapped the action in policy with better Q(s,a').Continued further till o better policy other than current policy.

Linear Programming

Using the pulp library
Created n variables one for each state,And added nk constraints that we
V* >= V(S),for each action.Thus returns a V*,further optimal policy could be computed from argmax(Q(s,a)).

Observation

All three algorithms yielded similar results,but in Maze problem
Linear Programming was better of the remaining two algorithms,as todays solvers can handle too many variables.Howard's Policy iteration was too slow compared to value iteration.
Linear Programming is order of poly(n,k),where as HPI is exponential in no of actions.

MDP(encoder):

Converted the Maze file into a grid of 2d array,after iterating through the grid the ones which are not of value 1, are considered as states and assigned state numbers.
After obtaining all states looped over all the states,and append the all transitions from each states.
Considered a reward of -1 for each transition, and reward of 100000 for reaching the final states..Discount factor of 0.99.
Some other possibles that were tried,are of reward=-1/(no of states) and final reward of 1.Decoder was to transversal from start state till end state and appending all the action,returning the output.