CS 6300 **HW04: Value and Policy Iteration** **Due Feb 8**

1 Value Iteration

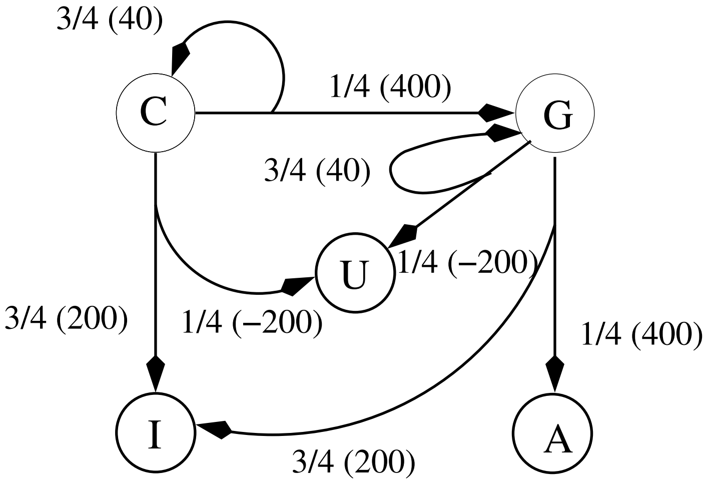
In the MDP below, there are 5 states: C(ollege), G(rad school), I(ndustry), A(cademia), and U(nemployed). States I, A and U are terminal states. Probabilities of transitions are either 1/4 or 3/4, and the values in parentheses are the rewards for that transition. The possible actions from states C and G are:

- State C: You may choose to stay in C, but with a probability of 1/4 you may end up going to state G.

  You may also choose to go to state I, but with probability 1/4 you end up in state U.

- State G: You may choose to stay in state G, but with probability 1/4 you end up in state U.

  You may also choose to go to state A, but with probability 3/4 you end up in state I.



1. You start in state C. Perform two iterations of value iteration, where you first compute the Q values and then take the maximum of the Q values. The discount is $\gamma = 1$.

2. Perform policy extraction after these two iterations to find $\pi^*(s)$. Please show all work.

2 Policy Iteration

Consider again the following MDP. There are 5 states: C(ollege), G(rad school), I(ndustry), A(cademia), and U(nemployed). States I, A and U are terminal states. Probabilities of transitions are either 1/4 or 3/4, and the values in parentheses are the rewards for that transition. The possible actions from states C and G are:

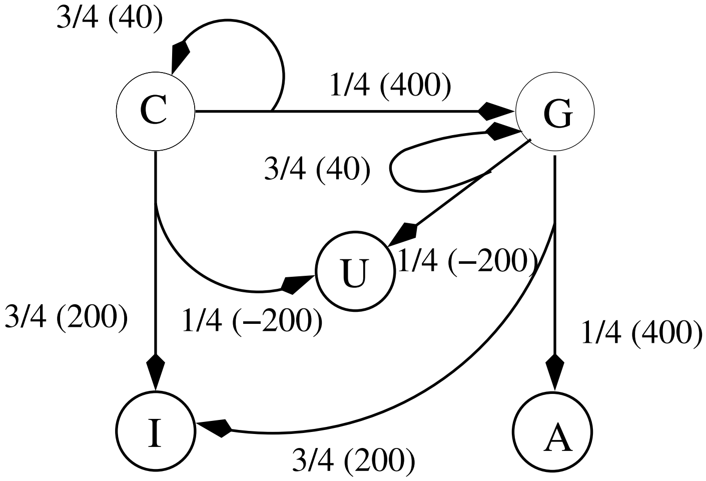
- State C: You may choose to stay s in C, but with a probability of 1/4 you may end up going g to state G.

  You may also choose to go g to state I, but with probability 1/4 you end up in state U.

- State G: You may choose to stay s in state G, but with probability 1/4 you end up in state U.

  You may also choose to go g to state A, but with probability 3/4 you end up in state I.

You start in state C. Assume your initial policy is $\pi_0(s) = s$, i.e., you wish to stay in the current state you're in. Also the discount is $\gamma = 1$.



For this problem you will perform one step of Policy Iteration:

1. Perform policy evaluation to solve for the utility values $V^{\pi_0}(C)$ and $V^{\pi_0}(G)$. Remember that the utility values can be solved for analytically.

   The analytical equation to use is:

$$V^{\pi_i}(s) = \sum_{s'} T(s, \pi(s), s') \left[R(s, \pi(s), s') + \gamma V^{\pi_i}(s')\right]$$

2. Perform policy improvement to find $\pi_1(s)$. Please show all work.