CPS 841 Reinforcement Learning Assignment 3

By: Daniel Bellissimo 500749419 and Tony Misic

1.)

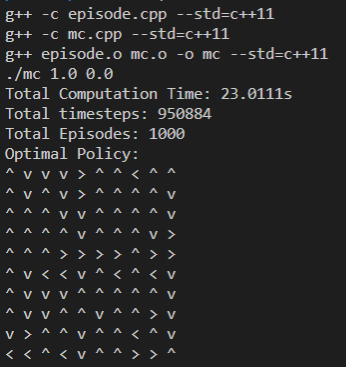
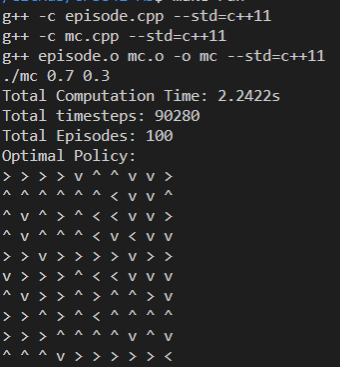
First Visit: By looking at the algorithm we can see that at the beginning we would go the last state, which would be a non-terminal state. Next, we set G = 0 and continue looping. During each iteration except the last, we continuously add 1 to the sum of rewards, by the last state we are at a total of 10. This is because we end on a non-terminal state. Therefore if we took the average of the Returns we would get 10 / 1, which equal 10, which is the value.

Every Visit: Ever visit follows the same principle expect you take the average for each state iteration not just the first case of the state. This means that we get a summation of n from 0 to 10 which equals 55 (n(n + 1) / 2). This divided by 10 (which is how many visits we do) equal to 5.5.

2.)

MC control seems to perform the slowest of all algorithms since it considers episodes of random length and random actions. Computation time is very unstable while never getting to the most optimal policy.

Here are some examples of first visit Monte Carlo control output.



To run the MC control policy:

g++ -c episode.cpp --std=c++11

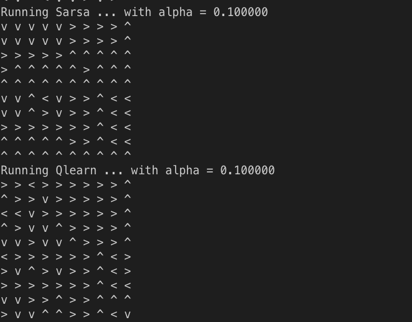
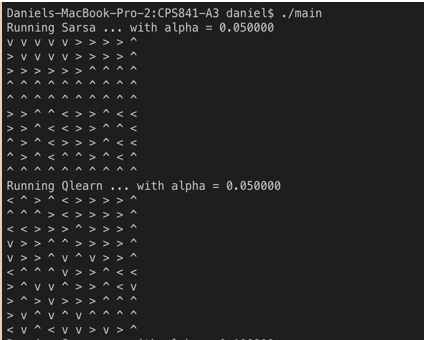
g++ -c mc.cpp --std=c++11

g++ episode.o mc.o -o mc --std=c++11

./mc 1.0 0.0

Q-Learning performs much better than Sarsa in terms of computation time and the resulting policy. Although Sarsa does converge to an optimal policy it does so after a much longer computation time due to the fact that the agent can get lost in endless loops while traversing the environment much easier as opposed to Q-learning where we just calculate the optimal state value action function and then use that as our policy.

Here are some examples of Sarsa vs Q learning:



To run the Sarsa and Q learning files use either

./run.sh

./main p1 p2

To compile the files and then execute the main file OR

Just execute the following:

g++ -c episode.cpp

g++ -c main.cpp

g++ -c sarsa.cpp

g++ -c q.cpp

g++ episode.o main.o sarsa.o q.o -o main

and THEN ./main p1 p2