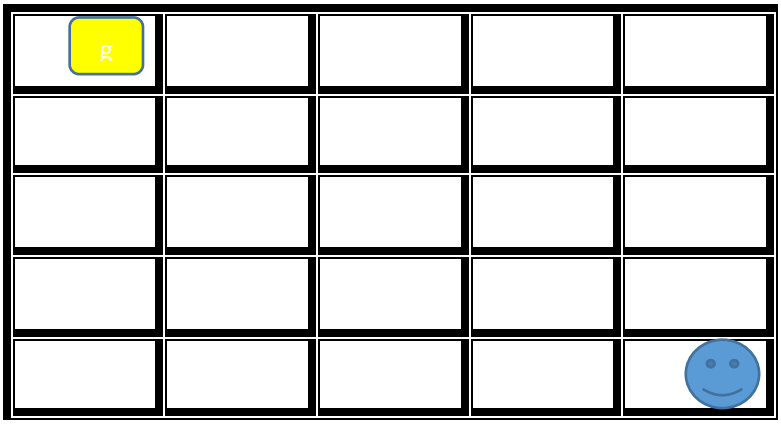
Experiments

We have a 5*5 Grid World, in which we are going to find our path to a goal. At first, we don't put any obstacles in the World and we get:

Table 1-Path finding grid world



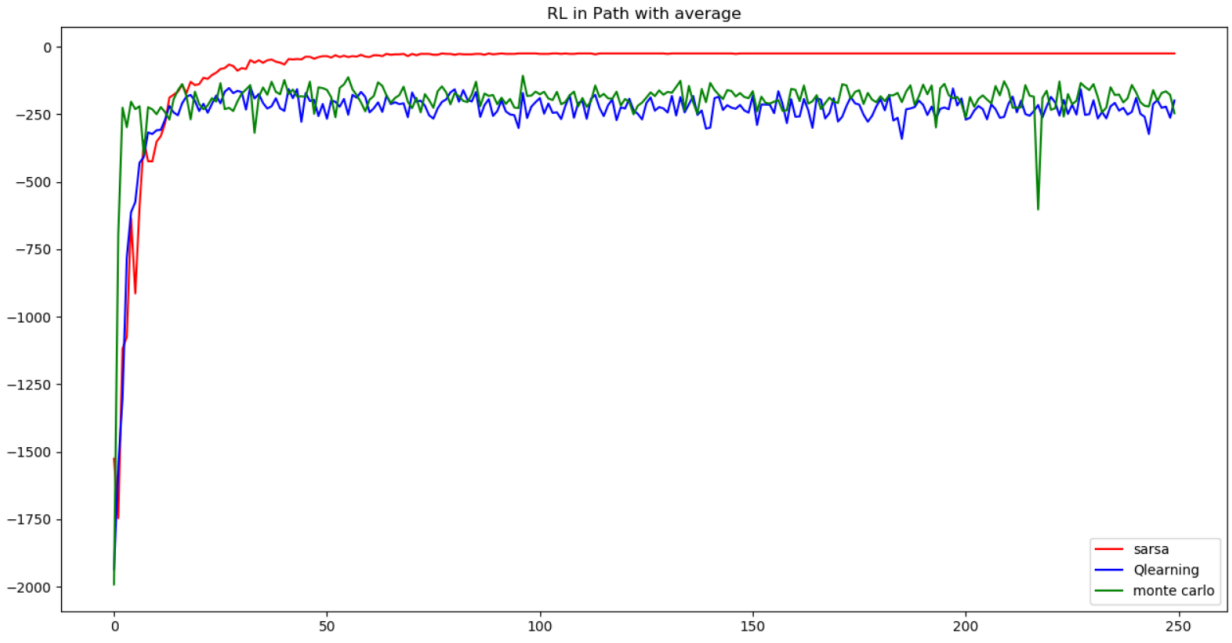


Figure 1-Path finding Comparison, Sampling 40, with average of return, over 30 runs

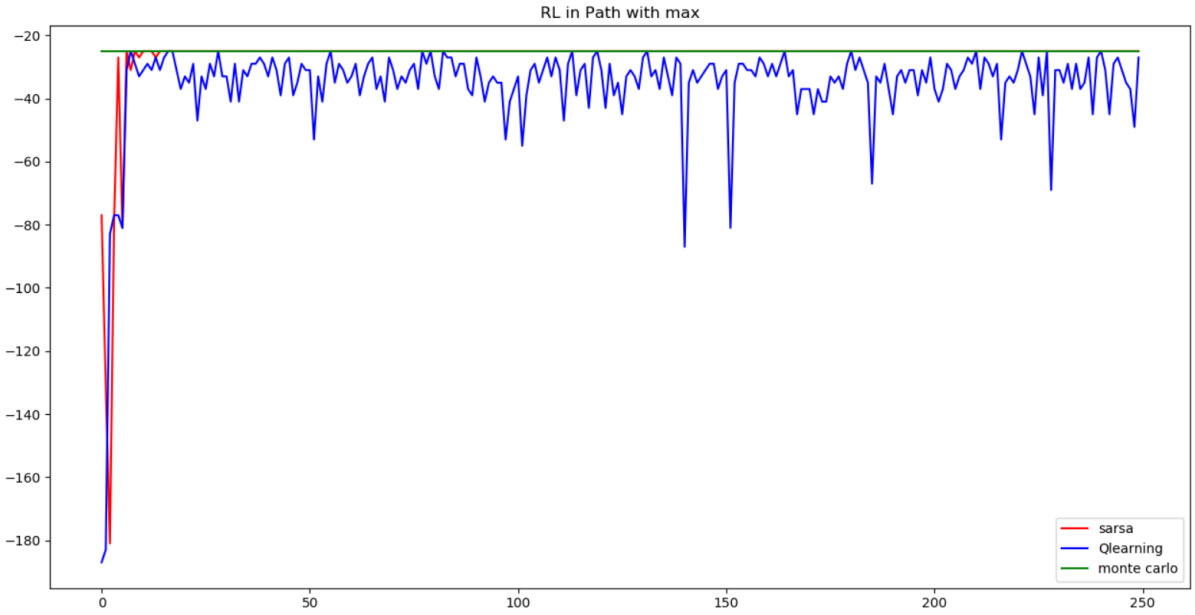
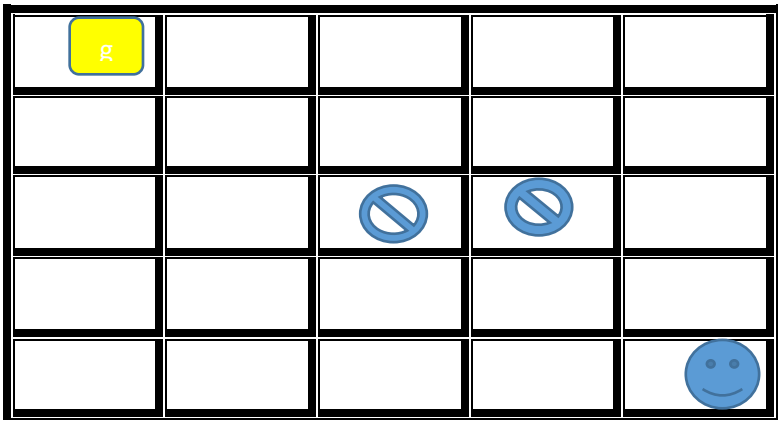


Figure 2-Path finding Comparison, Sampling 40, with maximum of return, over 30 runs

With these results we can draw a few conclusions, which are: Sarsa takes a while to be trained, however it will get better results and safer results in average. Because of the size of the problem Monte Carlo finds the best path in one of his runs, but it takes a while for other two algorithm to catch up.

After that we only put holes in the way of our agent. In this way we get these results:

Table 2-Path finding grid world with two obstacles



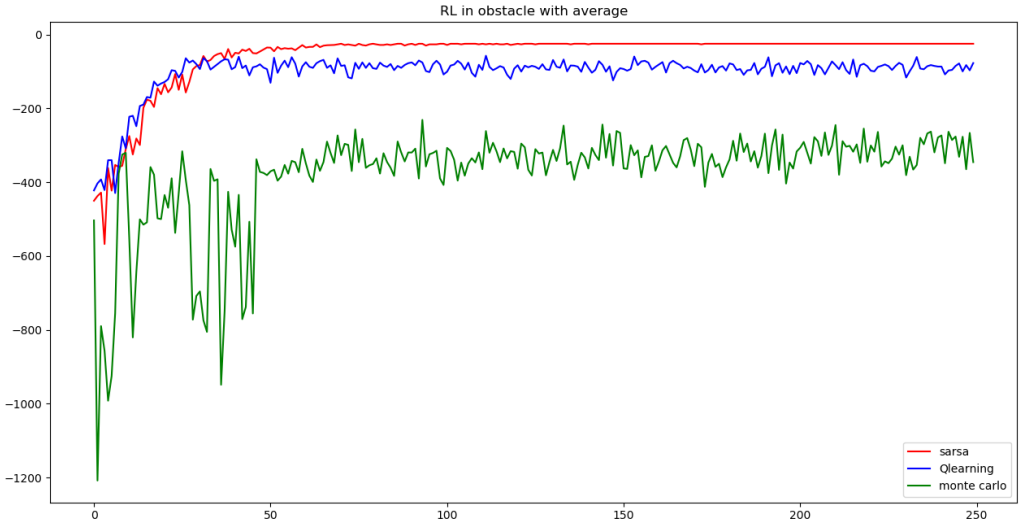


Figure 3-Obstacle Comparison, Sampling 40, with average of return, over 30 runs

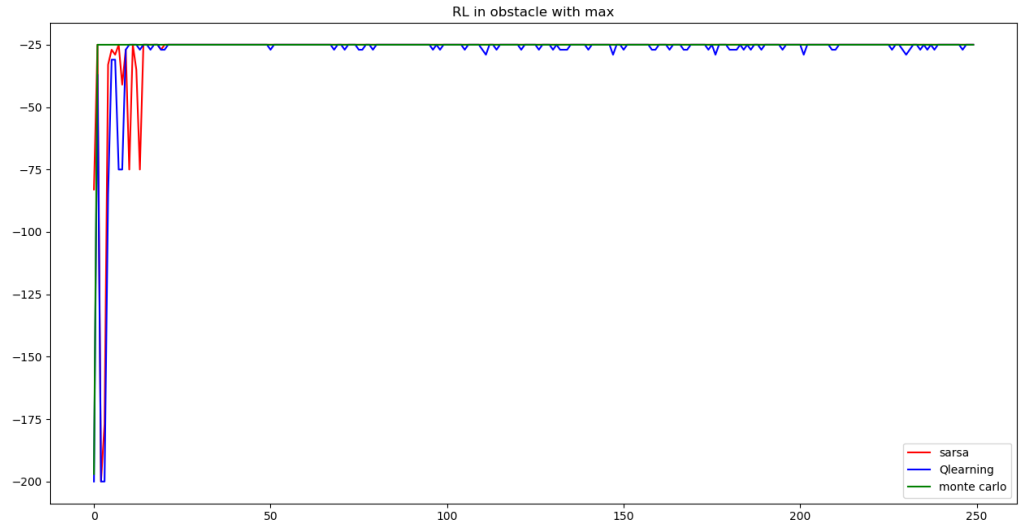
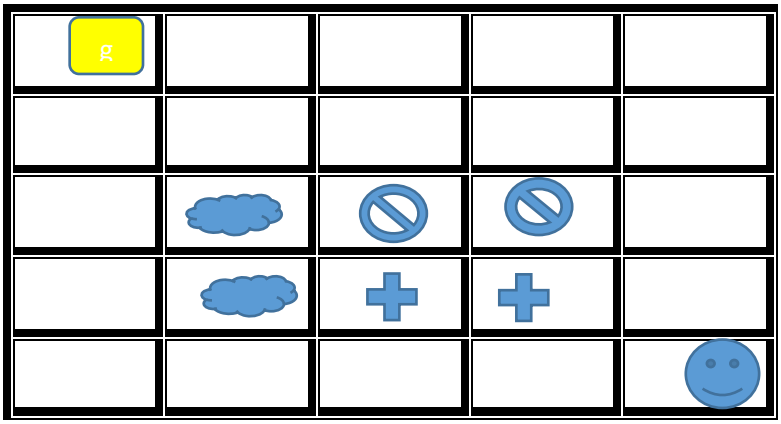


Figure 4-Obstacle Comparison, Sampling 40, with max of return, over 30 runs

With these results we conclude following. With obstacles in the path of our agent, Monte Carlo couldn't find the path as easy as before, however it still finds the way faster. Also, Sarsa again tries to find the best path possible while q learning is overestimating the value of some states, therefore in general it performs a little worse than Sarsa.

Finally, for a 5*5 grid we put not only Wampuses, but also some ammo so our agent could defend himself while he has to face more dangers. After these changes, our runs results in:

Table 3-game of wampus in grid world with two ammo and two wampus



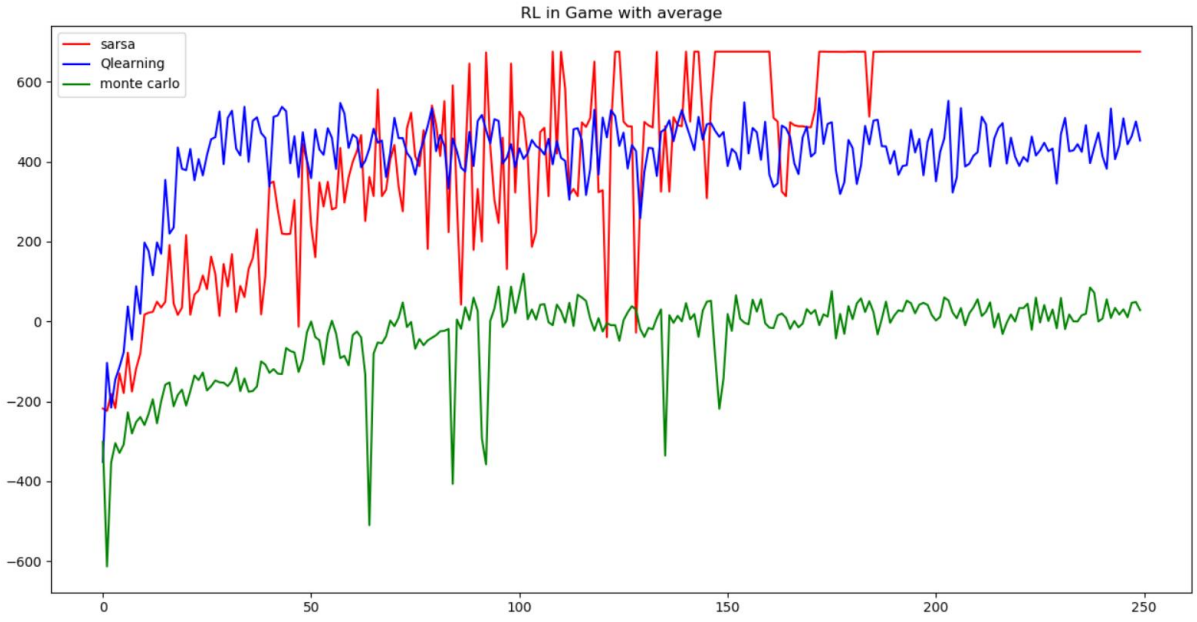


Figure 5-Game Comparison, Sampling 40, with average of return, over 30 runs

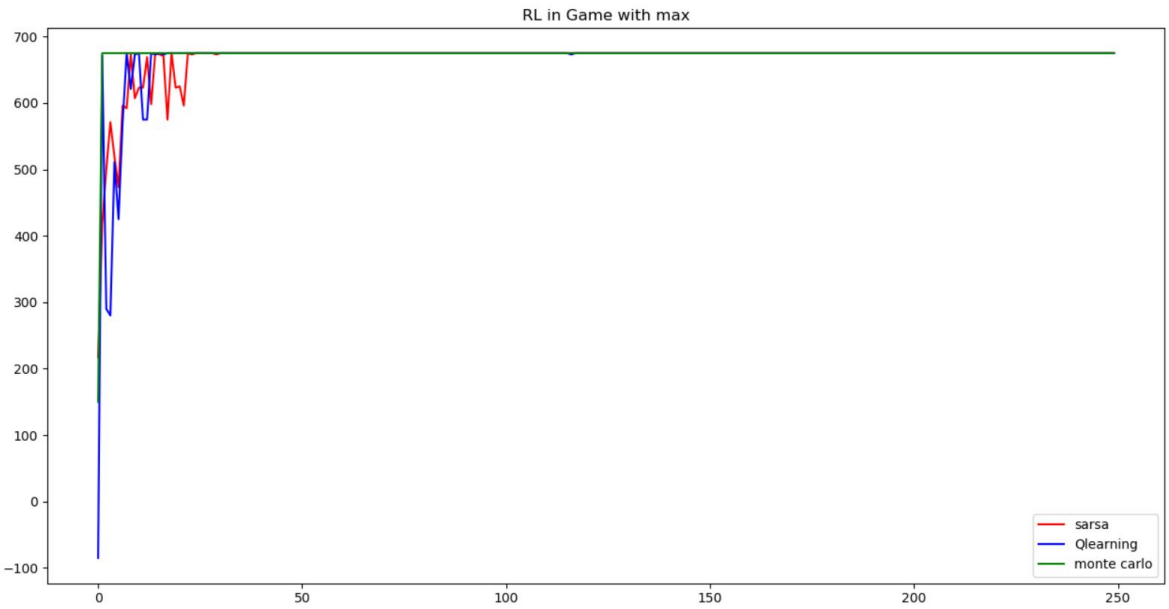


Figure 6-Game Comparison, Sampling 40, with maximum of return, over 30 runs

Which we can conclude, Monte Carlo because of pure randomness can't find the best way which is collecting ammo and killing monsters and finally going to goal most the times, but Sarsa most of time finds the best possible way to finishing the game. Also, as we can see once again that problem is getting more complicated q-learning is overestimating the value of action value function which at first it is good but in long run it cannot perform as well as Sarsa.

I also ran this problem for a 9*9 grid, however since the results and their conclusions were the same I did not put them here.

Future work

We can experiment when there is only 1 ammo and two Wampuses or we can make a maze and see whether our agent can find his way out or not.