MP 1 Part 2

Charles Shao
cshao4

September 29, 2014

1

a

Without random actions, the robot eventually settles on a non-optimal solution which steps on slippery tiles too often. This is probably because it does not explore options that are initially unrewarding, but become rewarding. It strictly follows what appears to be rewarding according to the Q-learning formula, which takes into account future tiles but discounts them.

b

With a small ϵ, a random action has a small probability. The value of rewards for episodes went up remarkably because the robot is finding paths that the Q-learning agent would deem to be unrewarding.

c

With a larger ϵ, the probability of doing a random action goes up even more. Even if the robot has already found the optimal action, it still might do something random, which wastes a step. Thus the reward went down from $\epsilon = 0.1$.

2

a

Episode rewards are poor and negative because the agent does not take into account the thief in its expected reward and calculation of the Q-value. (because the world does not tell the agent about the thief)

b

The rewards are much higher and the agent is successful. The world now tells the agent about the thief and his reward when bumping into thief from one state to another, so the agent can calculate smarter Q-values. The Q-values are dependant on the reliability of the world.

c

It seems that almost any value of rate and epsilon will yield similar results. My guess is that this is because the information coming from the world is very accurate and not misleading, and there are no "ridges" where reward is at first low and then very high.

3

The agent's reward is remarkably low at early episodes but increases and averages out as time goes on. This is because in early episodes, the agent still doesn't know much about the world. Averaging out ten runs gives a tighter bound on the performance of the agent and shows that the standard deviation is actually lower.

Figure 1: Ten simulations averaged.

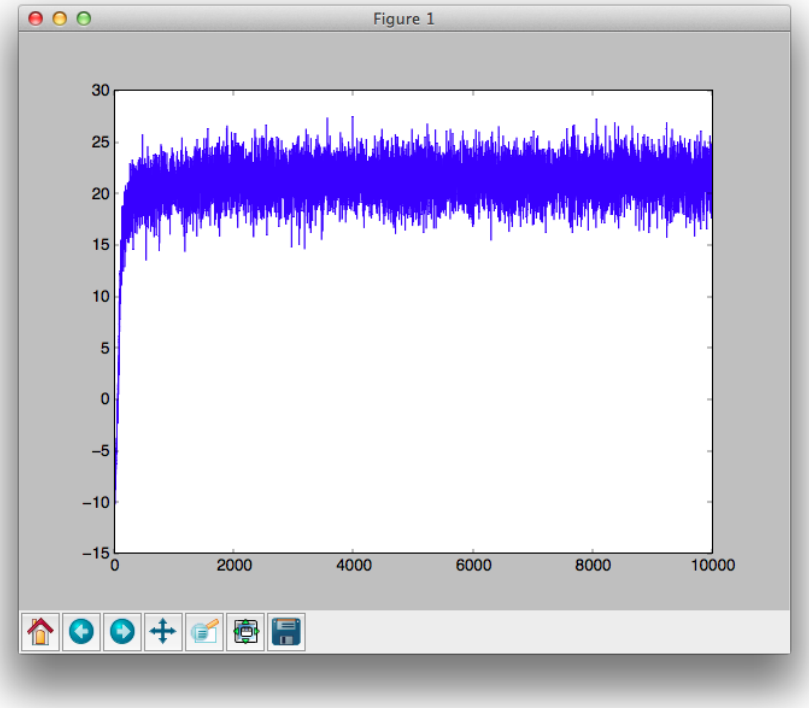

Figure 2: A single simulation.