CS 440

MP1 - Reinforcement Learning

September 29, 2014

Sam Laane <laane2@illinois.edu>

1 Part 2: Experimentation

1. **WorldWithoutThief**:

   (a) The agent is useless; it almost never moves anywhere. It never takes risks, and thinks its best bet is to stand still where it is safe.
   - Max expected discounted reward: 0.0
   - Min expected discounted reward: -4.0
   - Mean expected discounted reward: -0.004
   - Standard Deviation of expected discounted reward: 0.126427845034

   (b) The agent is now much better at reaching its goal. After a few Episodes, it start getting rewards. However, it is not very stable.
   - Max expected discounted reward: 256.0
   - Min expected discounted reward: -24.5
   - Mean expected discounted reward: 135.302
   - Standard Deviation of expected discounted reward: 53.754909506

   (c) At this point the agent starts doing worse. It almost always gets negative rewards.
   - Max expected discounted reward: 17.5
   - Min expected discounted reward: -77.0
   - Mean expected discounted reward: -29.1965
   - Standard Deviation of expected discounted reward: 14.9481148561

2. **WorldWithThief**

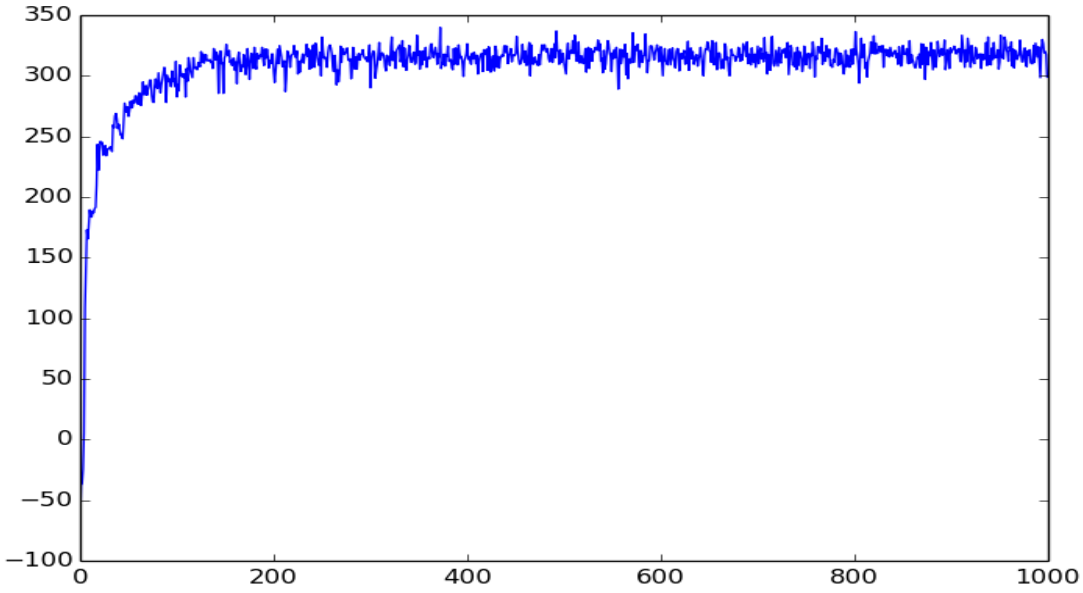
   (a) The agent is doing pretty poorly. It keeps bouncing between negative rewards.
   - Max expected discounted reward: -3.5
   - Min expected discounted reward: -32.5
   - Mean expected discounted reward: -12.842
   - Standard Deviation of expected discounted reward: 3.22467610777

   (b) The agent is doing pretty well now. Its score grows rapidly for about the first 50 Episodes. It has a large high score and is pretty stable.
   - Max expected discounted reward: 308.5
   - Min expected discounted reward: -58.0
   - Mean expected discounted reward: 273.8275
   - Standard Deviation of expected discounted reward: 29.2256906462

   (c) So to find the optimal learning rate, ϵ, I looked for the highest mean with the lowest standard deviation. I binary searched ϵ to the nearest hundredth place. I also graphed the learning rate to look for well-behaved behavior. I get the best behavior at about ϵ = .02, but it is hard to get more accurate due to the behavior being non-deterministic.
   - Max expected discounted reward: 340.0
   - Min expected discounted reward: -66.5
   - Mean expected discounted reward: 308.5615
   - Standard Deviation of expected discounted reward: 32.5212540925

   ‘

3. **Graphs**
   As you can see, it takes approximately it takes about 150 Episodes before it stabilizes. The variation is relatively small.



As predicted, this second graph(below) looks smother. However, it dose have a higher standard deviation for a reason I'm not sure of.

- Max expected discounted reward: 325.1
- Min expected discounted reward: -69.7
- Mean expected discounted reward: 306.63495
- Standard Deviation of expected discounted reward: 36.8329981266