ECE448/CS440 Wenyu Wu
Fall 2014 wenyuwu2
MP1_Part2

1.

(a) The agent move some steps and then stop completely in an corner.

The reason behind that is Going other ways(that is, leave the corner) has lower rewards than the block the agent is staying, thus the agent prefer crashing into the wall—i.e. stay at the same place without moving to avoid them.

(b)Sometimes the agent is able to deliver packages, go back and fetch new packages, and so on, and each time it's following the same route. Yet there are some cases that the agent completely stopped after deliver the first package.
Yet in general the agent works well, it is because when random activity is involved, the agent is able to break bad conditions with adventurous decisions.

(c)As epsilon goes bigger and bigger, random pattern started to appear in the agent's movement. Say, the agent will suddenly turn to some strange directions on its way of delivering. Because the bigger possibility of random choices, the more similar the agent is to a completely random agent. At this stage, packages are able to be delivered in most cases.

2.

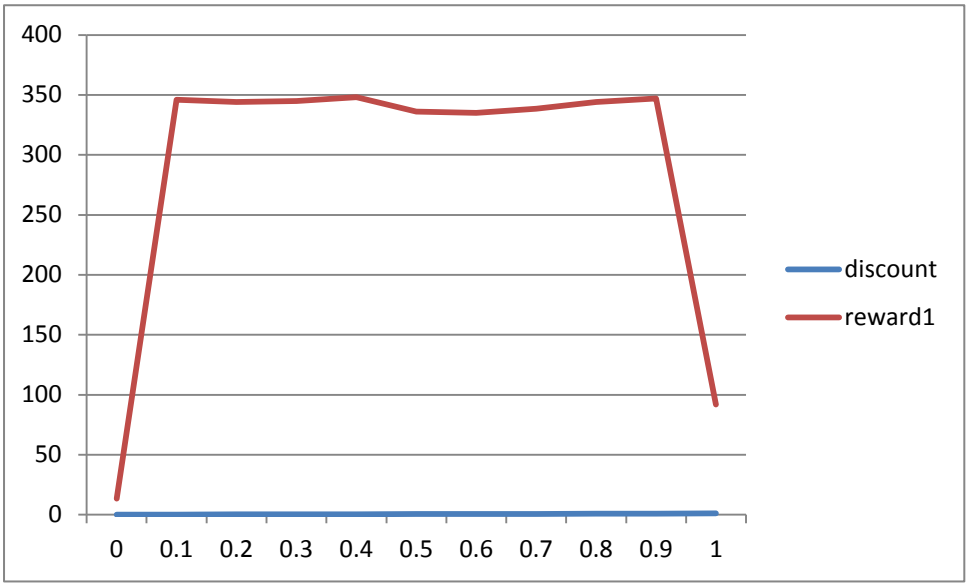
(a)The agents goes directly to a corner and remain stopped. Checking the episode file, I found that most rewards for each blocks are negative. This should be the reason, because whatever choice the agent makes, it will lose rewards, so the best choice is quickly find somewhere and hide inside that place.

(b)By setting the flag to y, the agent is doing a fantastic job: not only delivering packages successfully, but also able to evade the thief when necessary. Also rechecking the episode file, though at the beginning I see some negative reward number, but they turned positive later.

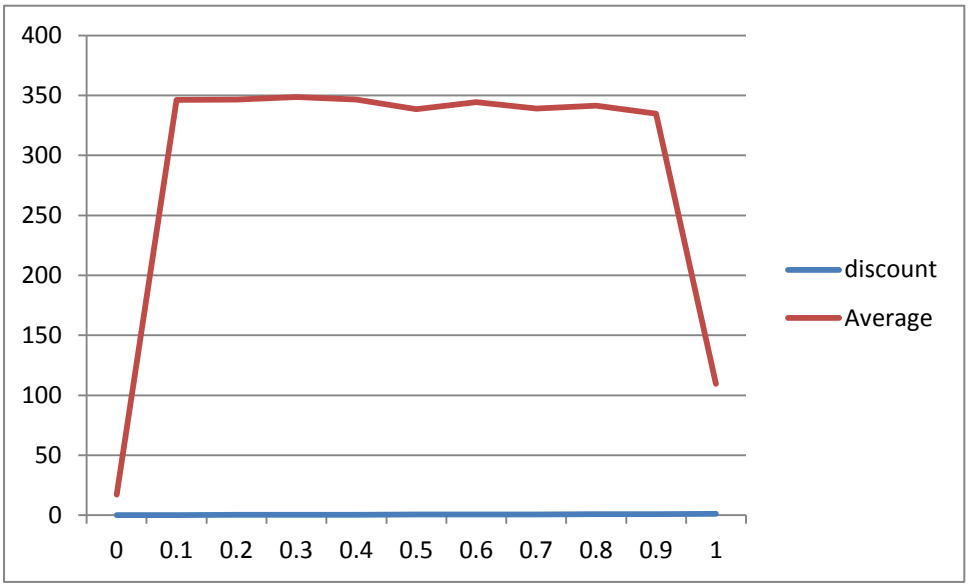
(c)The technique is keeping one value fixed, and play with another one. I found two combinations have quite good performance, the first one is rate = 0.2, epsilon=0.01; the second one is rate = 0.3, epsilon=0.001, also rate = 0.2, epsilon=0.005 gives a good result, the peak reward is around 330~350 for those three choices.
Best discount is around 0.2, and grows when epsilon drops, while epsilon value can vary a lot, it seems that 0.01 is an upper bound, and 0.001 is a lower bound. So we do need some random choices, but not a lot.

3.the value I use is
rate = 0.3, epsilon=0.001

dsct	rwd1	rwd2	rwd3	rwd 4	rwd 5	rwd 6	rwd7	rwd8	rwd9	rwd10	Average
0	13.5	15	16.5	17	20	17	15.5	19	18	21	17.25
0.1	346	352.5	326	354	346.5	343.5	345.5	348	349.5	351	346.25
0.2	344	346.5	344.5	351	350.5	347.5	346	352	352	332	346.6
0.3	345	348	342.5	351.5	343.5	347	354	345.5	351.5	357	348.55
0.4	348	350	327	354	345	350	351.5	347	342.5	350	346.5
0.5	336	353	349	334.5	335	353	332	326	322.5	345	338.6
0.6	335	350	344.5	350.5	354.5	342	342	344.5	351	330.5	344.45
0.7	338.5	346	359	331.5	333	340	325.5	342	337.5	336	338.9
0.8	344	335	344	338	334	351	347.5	341.5	348	330	341.3
0.9	347	327	326	341	331	335.5	333	330	345	332	334.75
1	92	94.5	80	123	99.5	143	127	133	121.5	82.5	109.6



And the average of ten experiements:



The table shows all the data comes from ten tests.

The graph for average data has a smoother curve.

Also the graph is in the shape of a platform, with sharp drop at two sides. And the average shows a slight drop of reward as discount grows.