Implementation of Reinforcement Learning Algorithms

Imanol Arrieta Ibarra

February 11, 2015

Because of the nature of the problem, performance can either be optimal or not. Optimal if after H time steps we observe a positive reword; not optimal, otherwise. So, to compute when the algorithm is typically achieving optimal performance we will compute local averages and report only those H's for which at 1000 episodes we achieve optimal performance more than 50% of the time. For local averages I refer to the average over 10 episodes.

1 UCRL

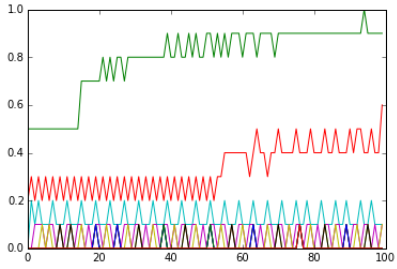

Figure 1: Average for each ten episodes.

As can be seen by figure 1, only $H = 1$(green line) and $H = 2$ (red line) achieve typical optimal performance after 1000 episodes.

2 PSRL

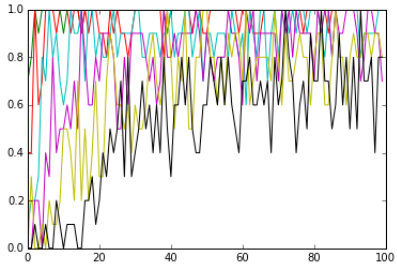

Figure 2: Average for each ten episodes for $H \in \{1, 9\}$.

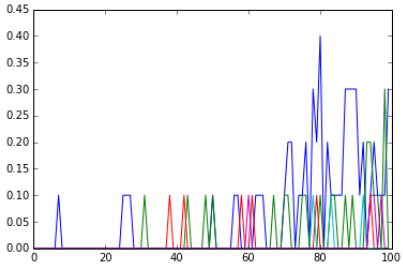

Figure 3: Average for each ten episodes for $H \in \{10, 15\}$.

Figure 2 shows the local averages for $H \in \{1, 9\}$. On the other hand 3 shows the local averages for H greater than 9. So the maximum H that typically attains optimum is 9.

3 Epsilon-PSRL

Figure 4 shows the local averages for $H \in \{1, 11\}$. On the other hand 5 shows the local averages for H greater than 11. So the maximum H that typically attains optimum is 11 with the proper epsilon-tuning.

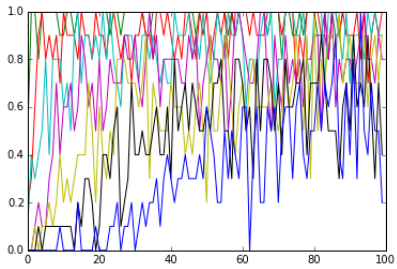

Figure 4: Average for each ten episodes for $H \in \{1, 9\}$.

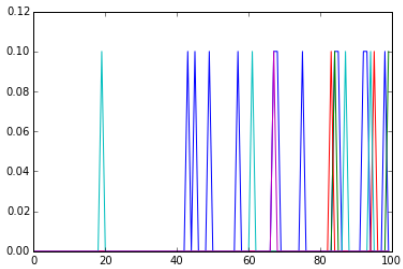

Figure 5: Average for each ten episodes for $H \in \{10, 15\}$.