Reinforcement Learning Project Report

Unsupervised and Reinforcement Learning in Neural Networks

Han JU, Hao REN

18 décembre 2012

1 Learning curve

2 Exploration & Exploitation

The parameter ϵ controls the balance between exploration and exploitation. When ϵ is big, the agent has more possibilities to take a random step in order to explore more place in the environement, while small ϵ makes the agent more likely to take a step according to the ϵ-greedy policy. The implementation of the algorithm SARAR depends on the choice of the ϵ, we have to make a tradeoff. If the ϵ is too big, the agent will seldom take a step with the greedy policy, the whole learning process is more stochastic, so the learning result is not significant. However, if ϵ is too small, the agent will be more likely to exploit the state around where it is, because the weights are not largely updated, the agent sometimes will travel around several same states, sometimes even not get to the goal within a given time step limit.

The best solution is to start with a big ϵ so as to explore the space as widely as possible. The ϵ decreases after each trails Thus, the first several trials with big ϵ will collect enough information about the weights (Q values), and the later trials with small ϵ can take a good use of these information via the greedy policy which makes the agent more likely to get into the goal area.

The learning curves and the average latency of the last 10 trails with ϵ=0.8, 0.6, 0.4, 0.2 are shown as below :

figure

$\epsilon = 0.8 \Rightarrow$ 219 steps

$\epsilon = 0.6 \Rightarrow$ 113 steps

$\epsilon = 0.4 \Rightarrow$ 77 steps

$\epsilon = 0.2 \Rightarrow$ 79 steps

3 Navigation Map

The naviagation map is shown as below :

//figure