Using Reinforcement Learning to solve the pendulum swing up

Bruno Costa

June 23, 2014

1 Objective

We want to solve the pendulum swing-up problem [Figure 1] using Reinforcement Learning techiniques, more specifically, using Actor-Critic framework. For that, we will consider both action and state space as continuos, and for such, we will need to use a Function Approximator (FA) for them both. The FA we will use is the tile coding. For the sake of organization, we will split our goal in two:

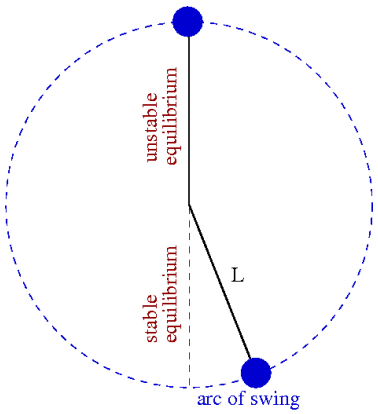

Figure 1: The Pendulum Swing-up problem

Solve the pendulum balancing
In this scenario, the pendulum start on top, and our policy must only balancing it there.

Swing-up
Finally, in this scenario, we will consider the full problem, starting the pendulum on the bottom.

Actor	
α	0.005
Critic	
α	0.1
Parameter	
γ	0.97
λ	0.65
σ	0.2
α-decay	0.75

Table 1: Parameters used in Pendulum Balancing

2 Pendulum Balancing

For the balancing, we changed the code and created a new environment. Now we have *env_mops_sim_up* and the first observation is $[\pi, 0]$. Using the attributes listed in Table 1 we could make it as we can see in Figures 2, 3 and 4.

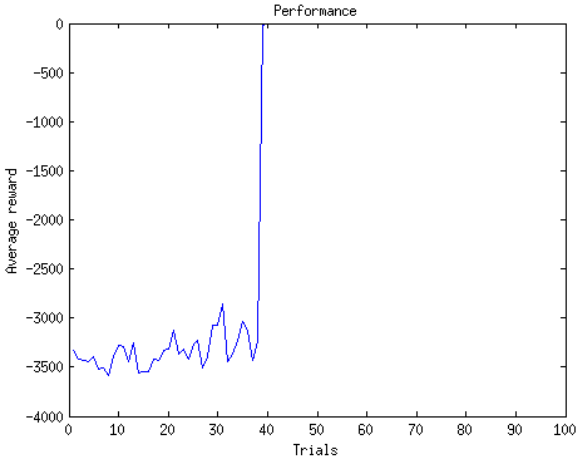

Figure 2: Total reward for each trial in the balacing problem

As we can see, it works(Figure 5). In Figure 2 we can see that the average reward is good after 40 trials, and as we low the α learning rate after that, it stays high. Otherwise, it diverges! Another interesting thing to notice is the policy in Figure 3: as the angle goes to the left, the action is to push to the right and vice-versa. The critic, in Figure 4 shows that the central position is the best.

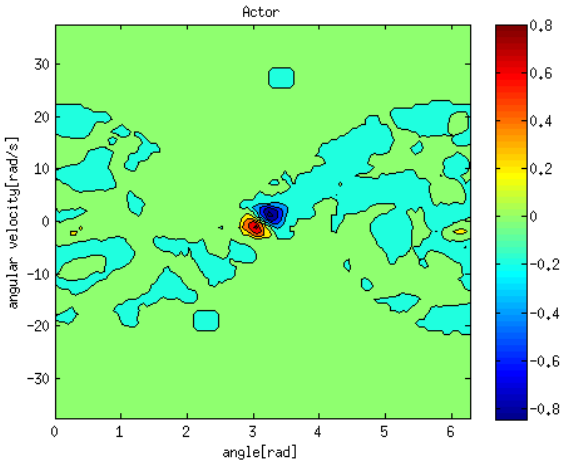

Figure 3: The Actor for the balancing problem

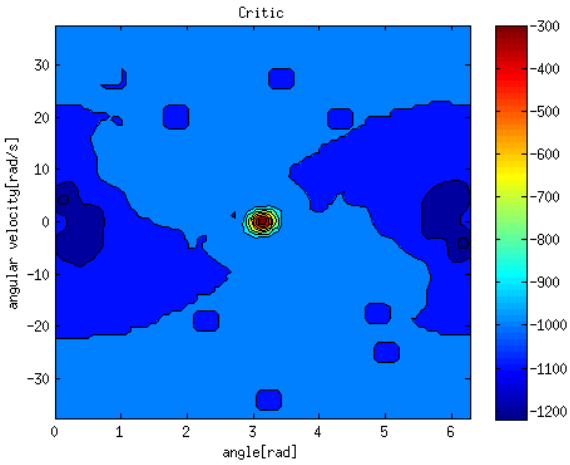

Figure 4: The Critic for the balancing problem

3 Sample Calculation

4 Results and Conclusions

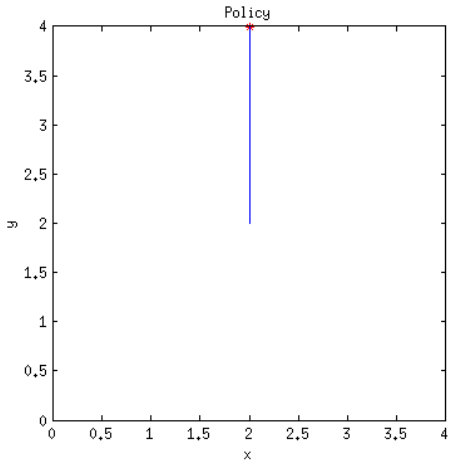

Figure 5: The final position for the balancing problem