N-step Actor Critic Algorithm Analysis

Exam id: B156424

1. Hypothesis about the effect of different N on algorithm performance

N step indicates that the value function of one state is updated according to the further N states. When N is 1, the value function of each state is very biased because it only calculates the reward and the discounted state value of the next state. Frequently Updating the networks will result in a longer learning time, but a good return can be obtained since it can more easily find the locally optimal returns. When N is large, the actor and critic loss functions for each update are large, so are their gradients. Therefore, the learning speed is faster.

2. Empirical proof

Figure1 shows the performance of the N-step Actor-Critic applying to the CartPole case. The AC1(1-Step Actor Critic) learns slower than other algorithms with bigger steps but it gains the highest averaged returns. The bigger the N is, the faster the algorithm learns.

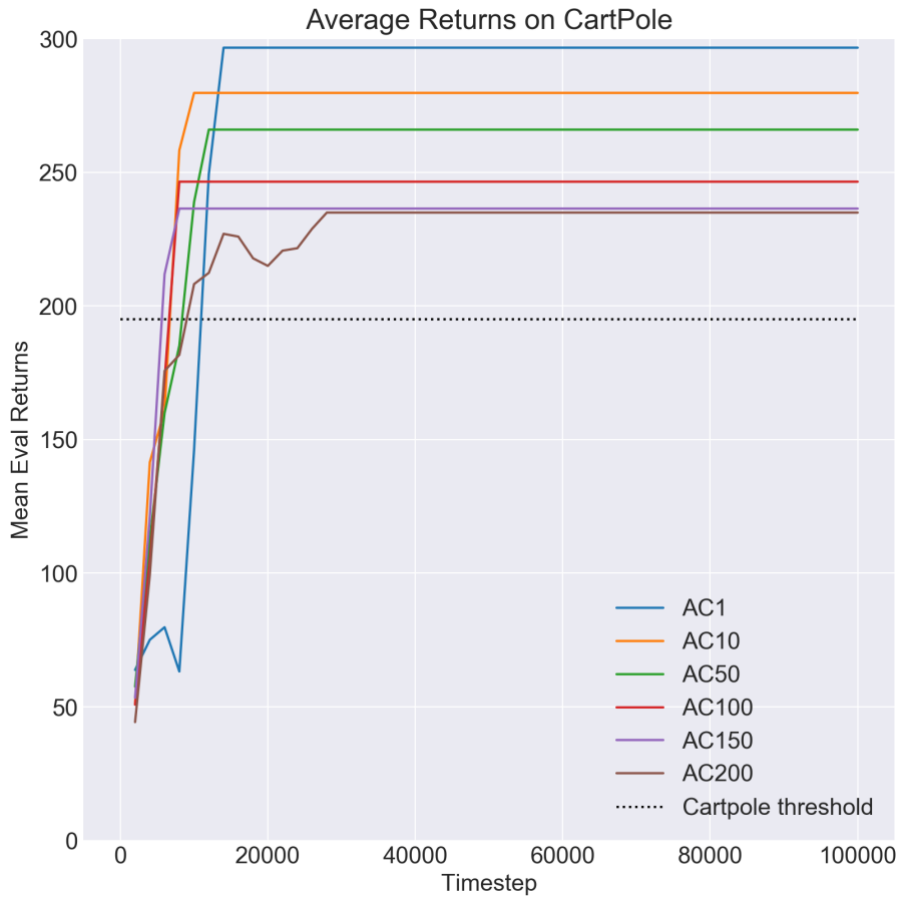

Figure 1: Train CartPole using1,10,50,150 and 200-step Actor Critic methods separately