

购买VIP年会员，独享全年GPU！目前限时特价，¥2999包揽全年课程和全年GPU，戳我速抢！(https://www.julyedu.com/sale/vip)

×

(https://www.julyedu.com)

登录

首页 (https://www.julyedu.com) 精品课程 (https://www.julyedu.com/category/index) 微专业 (https://www.julyedu.com/SmartSpec/index) (http://passport.julyedu.com/login) 问答社区 (https://ask.julyedu.com) / 全部问题 (https://ask.julyedu.com/explorer) / 深度学习 (https://ask.julyedu.com/explore/category-13) / 强化学习（Reinforcement learning）中Actor-Critic算法该如何深入理解？ (https://ask.julyedu.com/question/7967)

公开课 (https://www.julyedu.com/video/index)

社区 (https://ask.julyedu.com)

redirect=https://ask.julyedu.com/question/7967

## 强化学习（Reinforcement learning）中Actor-Critic算法该如何深入理解？

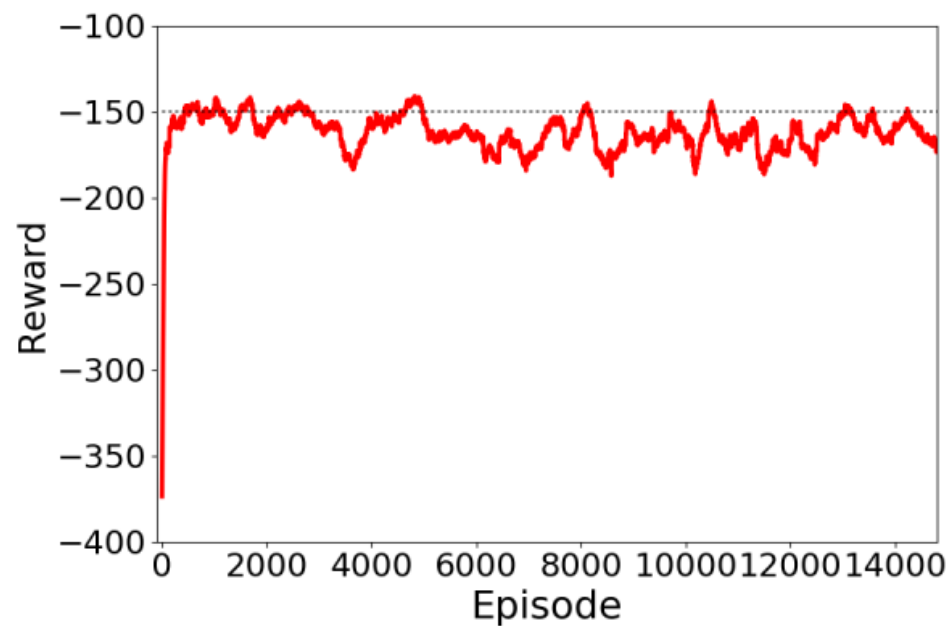
首先介绍一下OpenAI Gym的Pendulum问题。这是最简单的控制问题之一，适合新手来练习。

强化学习问题几乎都有state、action、reward这几个要素。

Pendulum问题是这样的：state是3维的，可以理解为钟摆的位置和速度action是1维的，在[-2, 2]之间，可以理解为你对钟摆用的力reward是个负数，越接近0越好具体描述参见 openai/gym控制钟摆的程序是这样的：初始化state观测到state，你自己的程序决定做出一个action环境（env）给你返回下一个state和你本次得到的reward循环2、3步

用神经网络举例；实际上你可以用线性函数、kernel等等方法做函数近似。Actor（玩家）：为了玩转这个游戏得到尽量高的reward，你需要实现一个函数：输入state，输出action，即上面的第2步。可以用神经网络来近似这个函数。剩下的任务就是如何训练神经网络，让它的表现更好（得更高的reward）。这个网络就被称为actorCritic（评委）：为了训练actor，你需要知道actor的表现到底怎么样，根据表现来决定对神经网络参数的调整。这就要用到强化学习中的“Q-value”。但Q-value也是一个未知的函数，所以也可以用神经网络来近似。这个网络被称为critic。

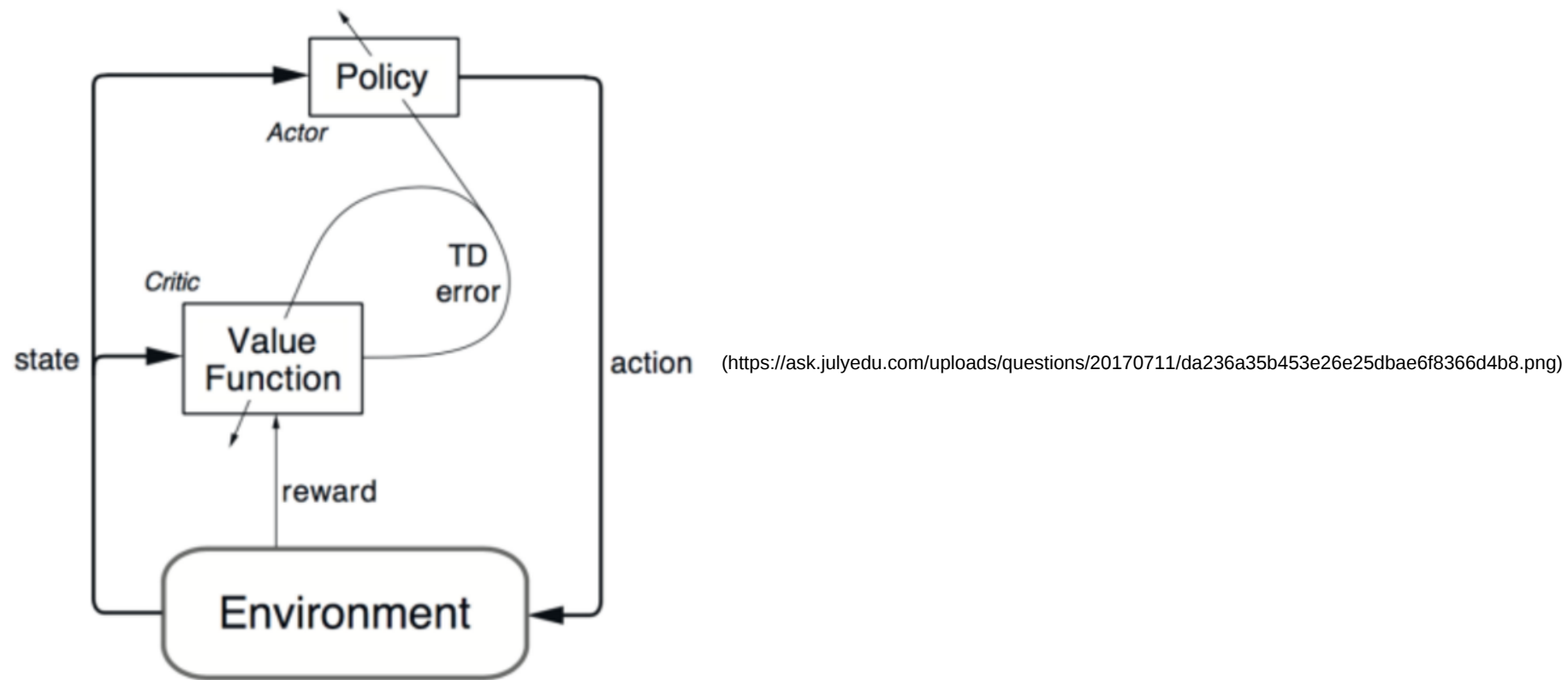


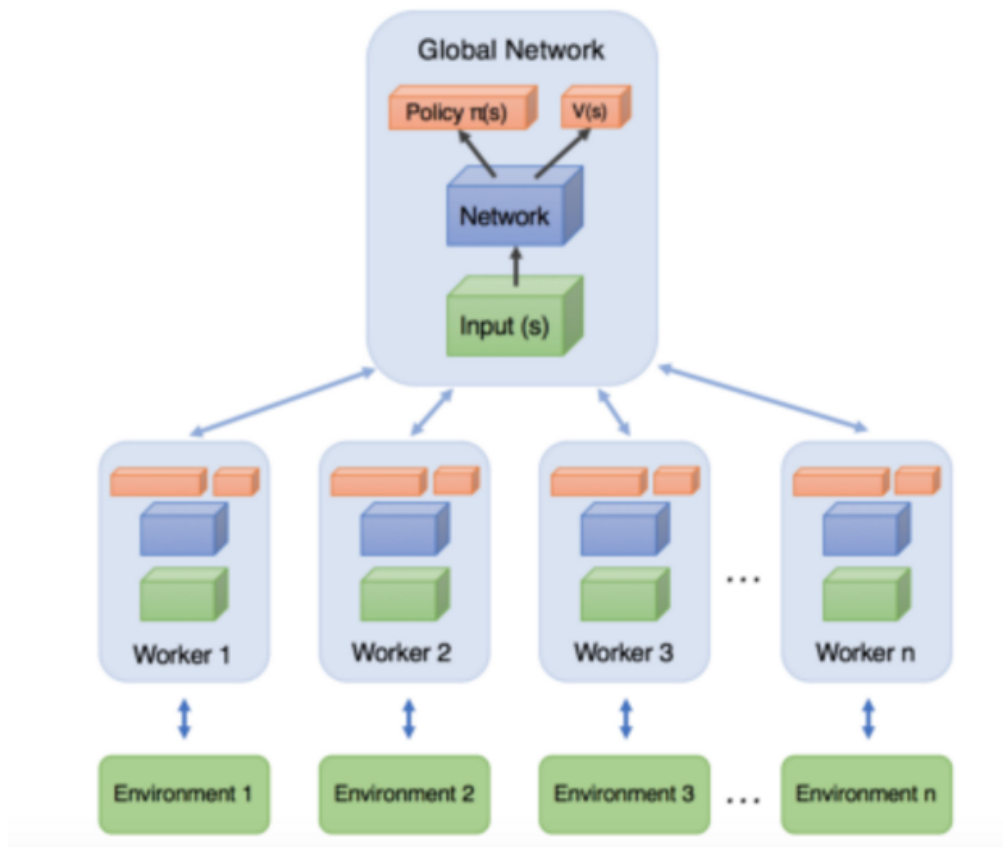


(<https://ask.julyedu.com/uploads/questions/20170711/b031f66ed01207ffe3919d7a9cf1cdc1.png>)

Actor-Critic的训练。我先通俗解释一下。Actor看到游戏目前的state，做出一个action。Critic根据state和action两者，对actor刚才的表现打一个分数。Actor依据critic（评委）的打分，调整自己的策略（actor神经网络参数），争取下次做得更好。Critic根据系统给出的reward（相当于ground truth）和其他评委的打分（critic target）来调整自己的打分策略（critic神经网络参数）。一开始actor随机表演，critic随机打分。但是由于reward的存在，critic评分越来越准，actor表现越来越好。







(<https://ask.julyedu.com/uploads/questions/20170711/37f55560e099070ef316f7930b58113a.png>)

2017-07-11 添加评论

分享

2 个回答



July (<https://ask.julyedu.com/people/July>) - 抠细节抠体验，不妥协不将就。

为了帮助大伙更好的入门和实战强化学习，七月在线也即将开《强化学习》实战课程拉  
(<https://ask.julyedu.com/people/July>)  
<https://www.julyedu.com/course/getDetail/85> (<https://www.julyedu.com/course/getDetail/85>)



0



0



2017-07-12

要回复问题请先登录 (<https://passport.julyedu.com/login?redirect=http%3A%2F%2Fask.julyedu.com%2Fquestion%2F7967>)或注册 (<https://passport.julyedu.com/reg>)

 (<http://weibo.com/askjulyedu>) 

[关于我们 \(https://www.julyedu.com/help/index/about\)](https://www.julyedu.com/help/index/about) · [联系我们 \(https://www.julyedu.com/help/index/contact\)](https://www.julyedu.com/help/index/contact) · [加入我们 \(https://www.julyedu.com/help/index/join\)](https://www.julyedu.com/help/index/join) · [意见反馈 \(http://ask.julyedu.com/question/55\)](http://ask.julyedu.com/question/55) · [刷题](#)

APP (<http://www.julyapp.com>)

七月在线科技©2016 · 京ICP备15036064号-2 (<http://www.miitbeian.gov.cn/>)

