**强化学习理论学习与代码实现**

# 深度强化学习-Prioritized Replay DQN

在Double DQN (DDQN)中，我们讲到了DDQN使用两个Q网络，用当前Q网络计算最大Q值对应的动作，用目标Q网络计算这个最大动作对应的目标Q值，进而消除贪婪法带来的偏差。今天我们在DDQN的基础上，对经验回放部分的逻辑做优化。对应的算法是Prioritized Replay DQN。

## 经验回放随机采样存在的问题

在Prioritized Replay DQN之前，我们已经讨论了很多种DQN，比如Nature DQN， DDQN等，他们都是通过经验回放来采样，进而做目标Q值的计算的。在采样的时候，我们是一视同仁，在经验回放池里面的所有样本都有相同的被采样到的概率。

但是注意到在经验回放池里面的不同样本由于TD误差的不同，对我们反向传播的作用是不一样的。TD误差越大，那么对我们反向传播的作用越大。而TD误差小的样本，由于TD误差小，对反向梯度的计算影响不大。在Q网络中，TD误差就是目标Q网络计算的目标Q值和当前Q网络计算的Q值之间的差距。

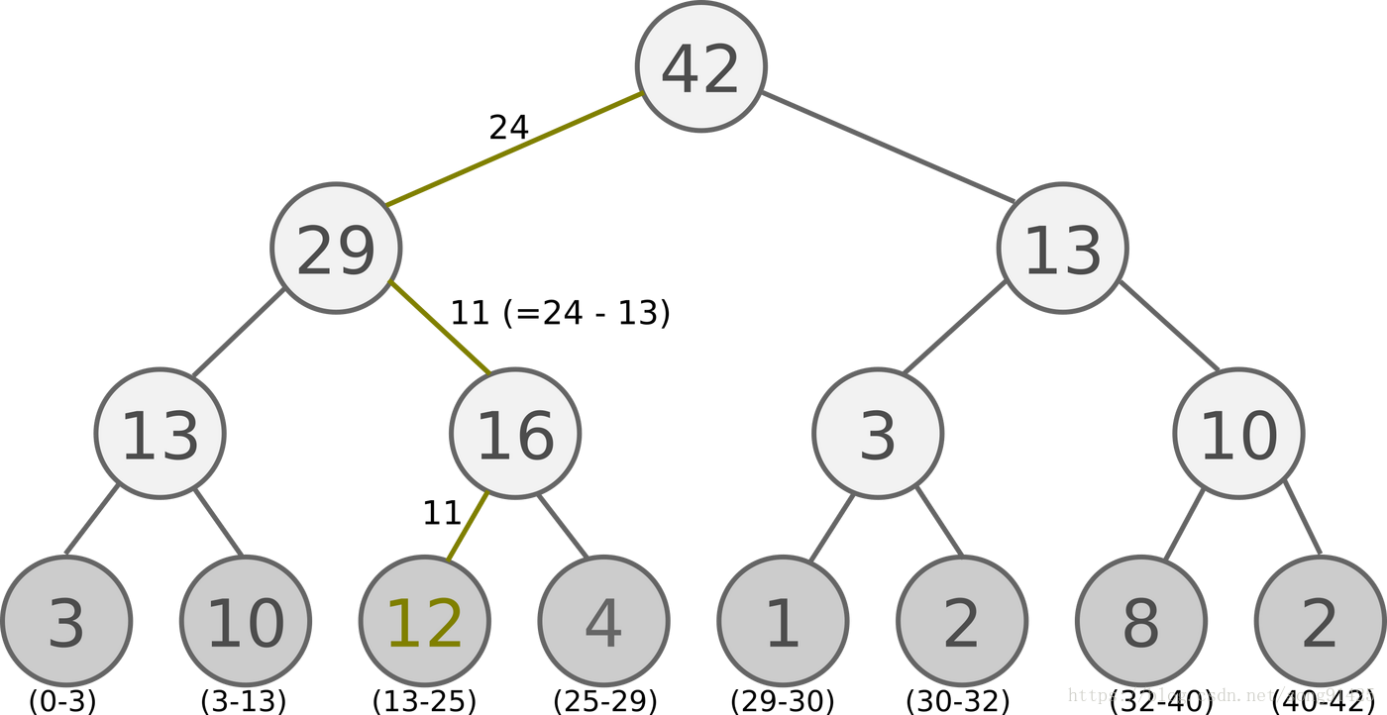
这样如果TD误差的绝对值|δ(t)|较大的样本更容易被采样，则我们的算法会比较容易收敛。下面我们看看Prioritized Replay DQN的算法思路。

## Prioritized Replay DQN算法的建模

Prioritized Replay DQN根据每个样本的TD误差绝对值|δ(t)|，给定该样本的优先级正比于|δ(t)|，将这个优先级的值存入经验回放池。回忆下之前的DQN算法，我们仅仅只保存和环境交互得到的样本状态，动作，奖励等数据，没有优先级这个说法。

由于引入了经验回放的优先级，那么Prioritized Replay DQN的经验回放池和之前的其他DQN算法的经验回放池就不一样了。因为这个优先级大小会影响它被采样的概率。在实际使用中，我们通常使用SumTree这样的二叉树结构来做我们的带优先级的经验回放池样本的存储。ptotal

具体的SumTree树结构如下图：



[1]

1. Schaul, T., et al. *Prioritized Experience Replay*. arXiv e-prints, 2016.